10-1-2012

Two and three dimensional segmentation of multimodal imagery

Sreenath Rao Vantaram

Two and Three Dimensional Segmentation of Multimodal Imagery

By

Sreenath Rao Vantaram

M. S. Electrical Engineering, Rochester Institute of Technology, Rochester, New York, United States of America, 2009

B.Tech, Electronics and Communications Engineering, Jawaharlal Nehru Technological University, Hyderabad, Andhra Pradesh, India, 2006

A dissertation submitted in partial fulfillment of the
requirements for the degree of Doctor of Philosophy in the
Chester F. Carlson Center for Imaging Science,
College of Science
Rochester Institute of Technology

October 2012

Signature of the Author _____

Accepted by _____
Dr. John P. Kerekes, Coordinator, Ph.D. Degree Program Date

CHESTER F. CARLSON CENTER FOR IMAGING SCIENCE
COLLEGE OF SCIENCE
ROCHESTER INSTITUTE OF TECHNOLOGY
ROCHESTER, NEW YORK

CERTIFICATE OF APPROVAL

Ph.D. DEGREE DISSERTATION

The Ph.D. Degree Dissertation of Sreenath Rao Vantaram has been examined and approved by the dissertation committee as satisfactory for the dissertation required for the Ph.D. degree in Imaging Science

_____
Dr. Eli Saber, *Dissertation Advisor*

_____
Dr. Sohail A. Dianat

_____
Dr. David W. Messinger

_____
Dr. Harvey Rhody

_____
Dr. Chance M. Glenn, Sr.

_____
Dr. Hossein Shahmohamad, *External Chair*

_____
Date

Dedication

This dissertation is dedicated to my family, friends, and colleagues, especially...

To my father, Mr. Viswanath Rao Vantaram, for his unparalleled and unique ways of motivating me from time to time, and inculcating in me the importance of education

To my mother, Mrs. Dharmavani Rao Vantaram, for her never-ending love

To my sister, Ms. Preethi Rao Vantaram, for being my unconditional source of support and encouragement

To my best friend Ms. Priyanka Yadlapati, for always being there for me

To all my dear friends in the RIT community Mr. Harsha Pairla, Ms. Divya Dinnepati, Mr. Yashwanth Gurram and the “408 club” for their unselfish companionship

To all my colleagues of the image, video and computer vision laboratory especially Dr. Mustafa Jaber, Mr. Abdul Syed, Mr. Siddharth Khullar, Mr. Yang Hu, Mr. Yilong Liang, Mr. Sankaranarayanan Piramanayagam, Mr. Osborn De Lima and Mr. Mohammad Yousefhussien, for their help and being a part of what has been a grueling yet enjoyable educational experience for me

Acknowledgements

I feel privileged and immensely happy to acknowledge the contributions and support of all the wonderful people who have been instrumental in the completion of my doctoral degree. The six years of graduate study at the Chester F. Carlson Center for Imaging Science and the Department of Electrical and Microelectronic Engineering, at the Rochester Institute of Technology (RIT), has taught me that a creative mindset, finest fellowman-ship and perceptiveness are the very essence of science and engineering. These attributes not only impart knowledge but also lay emphasis on the overall development of an individual, and I am immensely appreciative to the educational experience at RIT for ingraining these traits in me.

I would like to thank Dr. Sohail Dianat, Dr. David Messinger, Dr. Harvey Rhody, Dr. Chance Glenn, and Dr. Hossein Shahmohamad for their willingness to serve as committee members as well as their important individual contributions and supervision that have made this work possible. Also, I would like to thank Mr. Luis Garcia Ugarriza whose initial guidance was extremely valuable in helping me understand the field of segmentation.

Furthermore, I would like extend my sincerest gratitude to several people and funding organizations from industry. I am grateful to Mr. Mark Shaw from the Hewlett Packard Company and Dr. Vishwas Abhyankar from the DataPhysics Research Inc., for their generosity in supporting/sponsoring this research (in part) and showing tremendous satisfaction in my work and the results achieved. Additionally, I would like to offer my sincerest appreciation to the color/image science platform team from the Ricoh Production Print Solutions (RPPS LLC), comprising of Dr. Natalie Ross, Mr. Larry Ernst, Mr. Michael Stanich, Dr. Arianne Hinds, Mr. Katheek Chandu, and Mrs. Lori Cooper for not only providing me with multiple opportunities of being part of a high quality internship program, but also guiding me through it. Each of those experiences to me has been extremely fulfilling. Moreover, I would like to thank RPPS LLC for generously providing me with a constant source of funding in the form of a fellowship over the course of my doctorate which has always motivated me to work extremely hard towards achieving my academic goals.

Special thanks to Mrs. Sue Chan, Mrs. Patti Vicari, Mrs. Florence Layton, Mr. James Stefano, and Mrs. Joyce French for their academic assistance and support.

Finally, I am extremely indebted to the person whose unique ways of teaching, training and guidance is solely responsible for transmuting an ordinary or what I would like to call a “not-so-focused” student like me to a confident graduate researcher. His inspired notion that “if intuition and intelligence can find the right question, learning and research integrated with hard work and dedication help find the answer-seek and thou shall find”, has been instrumental in what I am today. Thank you, Dr. Eli Saber, for talking me into doing this doctorate and everything else that you have done. Although to get the work done given

by you it has taken me long work hours, innumerable sleepless nights/weekends, eating "off-the-refrigerator" food, not-so-frequent vacation, and 8 am presence in the lab (which I have not followed as much as you would have liked), it has been a steep educational experience being your student and I have enjoyed every bit of it.

Sreenath Rao Vantaram

Two and Three Dimensional Segmentation of Multimodal Imagery

By

Sreenath Rao Vantaram

Submitted to the Chester F. Carlson Center for Imaging Science in partial fulfillment of the requirements for the Doctor of Philosophy Degree at the Rochester Institute of Technology

Abstract

The role of segmentation in the realms of image understanding/analysis, computer vision, pattern recognition, remote sensing and medical imaging in recent years has been significantly augmented due to accelerated scientific advances made in the acquisition of image data. This low-level analysis protocol is critical to numerous applications, with the primary goal of expediting and improving the effectiveness of subsequent high-level operations by providing a condensed and pertinent representation of image information. In this research, we propose a novel unsupervised segmentation framework for facilitating meaningful segregation of 2-D/3-D image data across multiple modalities (color, remote-sensing and biomedical imaging) into non-overlapping partitions using several spatial-spectral attributes. Initially, our framework exploits the information obtained from detecting edges inherent in the data. To this effect, by using a vector gradient detection technique, pixels without edges are grouped and individually labeled to partition some initial portion of the input image content. Pixels that contain higher gradient densities are included by the dynamic generation of segments as the algorithm progresses to generate an initial region map. Subsequently, texture modeling is performed and the obtained gradient, texture and intensity information along with the aforementioned initial partition map are used to perform a multivariate refinement procedure, to fuse groups with similar characteristics yielding the final output segmentation.

Experimental results obtained in comparison to published/state-of the-art segmentation techniques for color as well as multi/hyperspectral imagery, demonstrate the advantages of the proposed method. Furthermore, for the purpose of achieving improved computational efficiency we propose an extension of the aforestated methodology in a multi-resolution framework, demonstrated on color images. Finally, this research also encompasses a 3-D extension of the aforementioned algorithm demonstrated on medical (Magnetic Resonance Imaging / Computed Tomography) volumes.

Table of Contents

Abstract ..... **vi**

Table of Contents ..... **vii**

List of Figures ..... **x**

List of Tables ..... **xvi**

Chapter 1: Introduction **1**

1.1 Problem Statement and Motivations 1

1.2 Research Goals 4

1.3 Original Contributions 6

1.4 Intellectual Merit and Broader Impacts 8

1.5 Publications and Intellectual Property representing Original Contributions 8

1.6 Organization 10

Chapter 2: Literature Review **12**

2.1 High Level Taxonomy 12

2.2 Low Level Taxonomy 13

2.2.1 Color/Gray Scale Image Segmentation Methodologies 13

2.2.1.1 Spatially Blind Approaches 13

2.2.1.1.1 Clustering 15

2.2.1.1.2 Histogram Thresholding 23

2.2.1.2 Spatially Guided Approaches 24

2.2.1.2.1 Region-Growing Approaches 26

2.2.1.2.2 Region-Merging Approaches 26

2.2.1.2.3 Hybrid Growing-Merging Approaches 27

2.2.1.2.4 Active Contours 29

2.2.1.2.5 Mumford-Shah Functional-based Approaches 33

2.2.1.2.6 Bayesian Segmentation Techniques 34

2.2.1.2.7 Graph-based Segmentation Techniques 38

2.2.1.2.8 Watershed-based Segmentation Techniques 43

2.2.2 Multi/Hyperspectral Image Segmentation Methodologies 46

2.2.3 3-D Segmentation Methodologies for Medical Volumes 49

2.3 Challenges with Conventional Techniques 51

Chapter 3: Automatic Color Image Segmentation by Dynamic Region Growth and Multivariate Region Merging **52**

3.1 Color Space Conversion 52
3.2 Edge detection for Color Imagery 53
3.3 Adaptive Threshold Generation 55
3.4 Initial Clustering 56
3.5 Region Growth and Dynamic Seed Addition (RGDSA) 57
3.6 Texture Characterization 63
3.7 Region Merging using a Multivariate Analysis 64
3.7.1 Multivariate ANalysis Of Variance (MANOVA) 65
3.7.2 Region Merging using MANOVA 68
3.8 Summary 69
Chapter 4: Multiresolution Adaptive and Progressive Gradient based Color Image Segmentation 70
4.1 Dyadic Wavelet Decomposition 71
4.2 Color Space Conversion and Gradient Computation 72
4.3 Adaptive Threshold Generation 73
4.4 Initial Clustering 74
4.5 Progressive Region Growth by Distributed Dynamic Seed Addition 75
4.6 Multi-resolution Seed Transfer 78
4.7 Texture Characterization and Multivariate Region Merging 80
4.8 Summary 82
Chapter 5: Automatic Spatial Segmentation of Multi/Hyperspectral Imagery by Fusion of Spectral-Gradient-Textural Attributes 83
5.1 Dimensionality Reduction 84
5.2 Multiband Gradient Detection for Multi/Hyperspectral Imagery 85
5.3 Initial Clustering and Region Growing 90
5.4 Co-occurrence Features based Texture Modeling 95
5.5 Multivariate Region Merging 97
Chapter 6: Synthesis of Intensity Gradient and Texture Information for Efficient Three-Dimensional Segmentation of Medical Volumes 99
6.1 3-D Vector Field Gradient Estimation 99
6.2 Initial Clustering and 3-D Volume Growing 102
6.3 3-D Entropy-based Texture Modeling 108
6.4 3-D Volume Merging using Analysis of Variance 108

6.5 Semi-Automatic 3-D Segmentation using the Proposed Framework 109
6.5.1 Sub-Volume of Interest (SVOI) Selection 109
6.5.2 Gradient Estimation and Volume Growing 110
Chapter 7: Experimental Results and Discussions 112
7.1 GSEG Algorithm Results 112
7.2 MAPGSEG Algorithm Results 119
7.3 Multi/Hyperspectral Image Segmentation Results 127
7.3.1 Qualitative Evaluation 127
7.3.2 Quantitative Evaluation: Segment-based Classification 133
7.4 3-D Segmentation Results of Medical Volumes 141
7.4.1 Qualitative Evaluation 142
7.4.2 Quantitative Evaluation 149
Chapter 8: Summary of Accomplishments, Conclusions and Potential Directions of Future Research 151
8.1 Summary of Accomplishments 151
8.2 Conclusions 152
8.3 Potential Directions of Future Research 152
8.3.1 Digital / Full Motion Video Segmentation 152
8.3.1.1 Proposed Approach for FMV Segmentation 153
8.3.1.2 Preliminary Results 154
8.3.2 Integration of Segmentation and Evaluation 156
Appendix A: Principle Component Analysis 158
Appendix B: Quantitative Evaluation of Segmentation Methods Using Multiple Ground Truths 159
B.1 Rand Index (R) 159
B.2 Probabilistic Rand (PR) Index 160
B.3 Normalized Probabilistic Rand (NPR) Index 160
Bibliography 162

List of Figures

1.1 (a) Color image, (b, c) show a matrix of region labels at two arbitrary locations in the image, (d) a complete pseudo colored segmentation result corresponding to the input color image. 2
1.2 Berkeley segmentation benchmark [Martin et al., 2001]: (a) Original images, and (b)-(f) region boundaries of multiple manually generated segmentations overlaid on the images. 3
1.3 A bird's eye view of research goals 5
2.1 High level taxonomy of image segmentation algorithms 12
2.2 Low level taxonomy of color/gray scale image segmentation algorithms 14
2.3 Sample color images with their corresponding 3-D point clouds 15
2.4 Results of the EDISON system [Christoudias et al., 2002] 18
2.5 Self-Organizing Map (SOM) in a rectangular neural layout 19
2.6 Results of the CTM algorithm [Yang et al., 2008] 22
2.7 Sample color images with their corresponding 3-D color histograms 23
2.8 Seed pixels (left), and region formed after a few iterations of growing (right) 25
2.9 Initial regions (left), and updated region map after a few iterations of merging (right) 25
2.10 Results of the JSEG algorithm [Deng et al., 2001] 28
2.11 Results of the DCGT algorithm [Balasubramanian et al., 2008] 28
2.12 Level set segmentation results obtained using an open source toolbox [Sumengen, 2005] 31
2.13 Results of the GRF-based segmentation algorithm in [Vantaram et al., 2011] 36
2.14 Graph theoretic formulation of an image 38
2.15 Results of the interactive graph cuts-based segmentation algorithm in [Boykov et al., 2001] acquired using an implementation provided by [Gulshan et al., 2010]. For each image shown is user defined foreground (white) and background (red) brush strokes as well as corresponding segmentation outcomes. 40
2.16 Results of the graph-based segmentation algorithm in [Felzenszwalb et al., 2004] 42
2.17 Sample image and 3-D topographic relief of its gradient 43
2.18 Results of the $gPb-owt-ucm$ algorithm [Arbeláez et al., 2011] 45
2.19 Results of the algorithm in [Hoang et al., 2005] 46
3.1 Block diagram of the proposed Gradient SEGmentation algorithm 52
3.2 Comparison of gradient information (RGB vs. CIE L*a*b*) of: (a) Cars, (b) Cheetah 53
3.3 Adaptive threshold generation scheme for Region Growth and Dynamic Seed Addition 56
3.4 (a) Input image I_{RGB}, (b) color converted image) I_{Lab}, (c) Gradient map G_{Lab}, (d) parent seeds (*PSs*) map, (e) final *PSs* map after employing size constraints 57

3.5 Flowchart of the region growth and dynamic seed addition process 58
3.6 (a) Input image I_{RGB}, (b) color converted image) I_{Lab}, (c) Gradient map G_{Lab}, (d) final *PSs* map after employing size constraints, (e) child seeds (*CSs*) generated on threshold increment, (f) parent seed boundaries $PS_{boundaries}$, (g) adjacent child seeds, (h) seed map at the end of 1st region growth interval, (i) spatially independent seeds generated during dynamic seed addition, (j) parent seed map for next stage of region growth. 59
3.7 Method to identify number of pixels per seed 60
3.8 (a) Gradient histogram of the 'Cars' image with adaptively generated segmentation thresholds for $1 \le n \le 5$ and $\Delta g = 10\%$, (b) seed map at the end of five *RGIs*, (c) neighborhood label assignment, (d) iterative morphological label assignment to yield final region growth map 62
3.9 Texture characterization: (a) I_{RGB}, (b) I_{Lab}, (c) indexed image, (d) texture channel 64
3.10 Region merging module (M4) using MANOVA 68
4.1 Overview of the MAPGSEG algorithm 70
4.2 (a) Two level decomposition (k=2) with corresponding designations, (b) Automatic determination of number of decomposition levels based on smallest desired dimension (d_{des}) 71
4.3 Comparison of gradient information using CIE $L^* a^* b^*$ data at different resolution levels of: (a) Cars, (b) Cheetah 72
4.4 Adaptive threshold generation 73
4.5 Initial Clustering phase for a 2-level decomposition: (a) Lowest resolution image I_{RGB2}, (b) I_{Lab2}, (c) G_{Lab2}, (d) *PSs* map, (e) *PSs* map after employing spatial constraints 74
4.6 Progressive Region Growth utilizing Distributed Dynamic Seed Addition 75
4.7 (a) *PSs* map, (b) small/isolated unsegmented regions enclosed within *PS*s, (c) direct label assignment of these isolated regions, (d) *PSs* map after seed saturation, (e) C*Ss* map, (f) boundaries of saturated *PSs* map, (g) Adjacent *CSs*, (h) seed map at the end of 1st RGI, (i) new seeds generated during dynamic seed addition, (j) *PS* map for 2nd RGI. 76
4.8 (a) Gradient histogram of the 'Cars' image with adaptively generated segmentation thresholds for $1 \le n \le 5$ and $\Delta g = 10\%$, (b) seed map at the end of five *RGIs*, (c) neighborhood label assignment, (d) iterative morphological label assignment to yield final region growth map 77
4.9 (a) Interim segmentation result at *L=k=2*, obtained at the end of processing in processing in M5 and M6, (b) acquired *HCSs* and *LCSs*, at *L=k-1=1 (j*th scale*)*, after the interim segmentation is processed in M4, (c) segregation of RGI utilizing G_{Labj}, $G_{Labj\text{-}HCSs}$, $G_{Labj\text{-}LCSs}$ 78

4.10 (a) Multi-resolution Seed Transfer module, (b) interim segmentation result (*L=k=2)*, (c) up-conversion (*L=k-1=1)*, (d) mode filtering, (e) gradient quantization, (f) seeds within the two lowest quantization levels, (g) *HCSs* acquired after post processing..... 79
4.11 Region merging module (M6) using MANOVA in the MAPGSEG algorithm..... 81
5.1 Bird's eye view of the proposed algorithm for multi/hyperspectral image segmentation 83
5.2 Proposed gradient computation scheme..... 86
5.3 (a) HYDICE 'Urban' scene (bands 49, 35, and 18), (b) Result using [Lee and Cok, 1991], and (c) Result using proposed gradient computation scheme 90
5.4 Initial clustering: (a) Initial parent seeds (*PSs*) map using g_{init}=2.5 %{G_{max}}, and (b) Spatially constrained *PSs* map using 20**MSS*=100pixels..... 91
5.5 Flowchart of the proposed region growing method 91
5.6 (a) Child seeds (*CSs*) map, (b) Parent seed boundaries ($PS_{boundaries}$), (c) Adjacent *CSs*, (d) Region map obtained after parent-child similarities have been processed, (e) Region map obtained at the end of a few iterations of the growth process, (f) New seeds non-adjacent to existent *PSs*, (g) *PSs* for next stage of region growth, and (h) Final region growth map 93
5.7 Gradient histogram of the HYDICE image shown in Fig. 5.3(a) divided into ten region growth intervals each spanning equal image area 95
5.8 (a) sample 4x4 image matrix, (b) computed co-occurrence matrix or gray tone spatial dependence frequency matrix (P) calculation using $\phi = 0^o$ and $\delta = 1$ 96
6.1 Overview of the proposed 3-D segmentation algorithm 99
6.2 3-D Gradient estimation: (a) Original MRI slices (#30 and #40), (b) corresponding edge maps, and (c) 3-D view (left) accompanied by a cross-sectional view (right)..... 102
6.3 Flowchart of the proposed volume growing method 103
6.4 Initial clustering: (a) Initial parent seeds (*PSs*) map using g_{init}=2.5 %{ G_{max}}, and (b) Spatially constrained *PSs* map using 50**MSV*=250 voxels. Volume growing: (c) Child seeds (*CSs*) map, (d) Parent seed surfaces ($PS_{surfaces}$), (e) Adjacent *CSs*, (f) Partition map obtained after parent-child similarities have been processed, (g) partition map obtained at the end of a few iterations of the growth process, (h) New seeds non-adjacent to existent *PSs*, (i) *PSs* for next stage of volume growing, and (j) Final volume growth map..... 104
6.5 (a) Gradient histogram of a sample MRI dataset (shown in Fig. 6.2(a)) divided into ten volume growth intervals each spanning equal area, and (b) portion of the gradient histogram boxed in green (displayed for viewing ease). 107
6.6 Overview of the proposed semi-automatic segmentation algorithm..... 109
6.7 Sub-Volume of Interest selection..... 110

7.1 (a) Original RGB image, (b) color space conversion, (c) gradient map, (d) seed map after region growth, (e) texture channel, and (f) final segmentation map after region merging..... 113
7.2 Moon light Results: (a) Original, (b) GRF, (c) JSEG, (d) EGS and (e) GSEG..... 113
7.3 Face Results: (a) Original, (b) GRF, (c) JSEG, (d) EGS, and (e) GSEG..... 114
7.4 Cheetah Results: (a) Original, (b) GRF, (c) JSEG, (d) EGS, and (e) GSEG 114
7.5 London Results: (a) Original, (b) GRF, (c) JSEG, (d) EGS, and (d) GSEG..... 115
7.6 Cars Results: (a) Original, (b) GRF, (c) JSEG, (d) EGS, and (e) GSEG 116
7.7 Distributional comparison of NPR scores for 300 images of the Berkeley database..... 117
7.8 Cumulative Percentage of NPR Scores..... 117
7.9 Parachute Results: (a) Original, (b) GRF, (c) JSEG, (d) EGS, and (e) GSEG..... 118
7.10 Mask Results: (a) Original, (b) GRF, (c) JSEG, (d) EGS, and (e) GSEG..... 118
7.11 Safari Results: (a) Original, (b) GRF, (c) JSEG, (d) EGS, and (e) GSEG..... 118
7.12 Fox Results: (a) Original, (b) GRF, (c) JSEG, (d) EGS, and (e) GSEG 118
7.13 Results of the MAPGSEG algorithm at various stages of processing: (a) original RGB 'Star Fish' image, (b) color converted 'Star Fish' image, (c) color gradient, (d) seeds maps at the end of progressive region growth, (e) entropy based texture maps, (f) interim and final segmentation outputs..... 119
7.14 Skydiver results: (a) original RGB image, (b) GRF, (c) JSEG, (d) EGS, and MAPGSEG (e) L=2, (f) L=1, (g) L=0 120
7.15 Church results: (a) original RGB image, (b) GRF, (c) JSEG, (d) EGS, and MAPGSEG (e) L=2, (f) L=1, (g) L=0..... 121
7.16 Boat results: (a) original RGB image, (b) GRF, (c) JSEG, (d) EGS, and MAPGSEG (e) L=2, (f) L=1, (g) L=0 122
7.17 Cheetah results: (a) original RGB image, (b) GRF, (c) JSEG, (d) EGS, and MAPGSEG (e) L=2, (f) L=1, (g) L=0 122
7.18 Race cars results: (a) original RGB image, (b) GRF, (c) JSEG, (d) EGS, and MAPGSEG (e) L=2, (f) L=1, (g) L=0 123
7.19 Distribution of NPR scores for 300 images of the Berkeley database 124
7.20 Cumulative Percentage of NPR Scores..... 124
7.21 Star fish results: (a) original RGB image, (b) GRF, (c) JSEG, (d) EGS, and MAPGSEG (e) L=2, (f) L=1, (g) L=0 125
7.22 Chapel results: (a) original RGB image, (b) GRF, (c) JSEG, (d) EGS, and MAPGSEG (e) L=2, (f) L=1, (g) L=0 126

7.23 Air force results: (a) original RGB image, (b) GRF, (c) JSEG, (d) EGS, and MAPGSEG (e) L=2, (f) L=1, (g) L=0 126
7.24 Dolphins results: (a) original RGB image, (b) GRF, (c) JSEG, (d) EGS, and MAPGSEG (e) L=2, (f) L=1, (g) L=0 126
7.25 (a) HYDICE 'Urban' scene (bands 49, 35, and 18), (b) Gradient map, (c) Region growth map, (d) f_{ASM}, (e) f_{CON}, (f) f_{ENT}. Segmentation results: (g, h, i) Proposed, (j, k, l) SCRM, and (m, n, o) MHIS 128
7.26 (a) HYDICE 'Washington' scene (bands 49, 35, and 18). Segmentation results: (b, c, d) SCRM, (e, f, g) MHIS, and (h, i, j) Proposed 130
7.27 (a) Hyperion 'San Francisco' scene (bands 40, 35, and 18). Segmentation results: (b, c, d) SCRM, (e, f, g) MHIS, and (h, i, j) Proposed 131
7.28 (a) MISI scene (bands 26, 15, and 5). Segmentation results: (b, c, d) SCRM, (e, f, g) MHIS, and (h, i, j) Proposed..... 132
7.29 (a) ROSIS 'University' scene (bands 60, 30, and 2). Segmentation results: (b, c, d) SCRM, (e, f, g) MHIS, and (h, i, j) Proposed 134
7.30 ROSIS 'University' scene: (a) Training, and (b) Testing samples..... 135
7.31 GML classification results: (a) Pixel-based, (b) SCRM, (c) MHIS, and (d) Proposed. SVM classification results: (e) Pixel-based, (f) SCRM, (g) MHIS, and (h) Proposed..... 136
7.32 (a) ROSIS 'Center' scene (bands 60, 30, and 2). Segmentation results: (b, c, d) SCRM, (e, f, g) MHIS, and (h, i, j) Proposed..... 137
7.33 ROSIS 'Center' scene: (a) Training, and (b) Testing samples..... 138
7.34 GML classification results: (a) Pixel-based, (b) SCRM, (c) MHIS, and (d) Proposed. SVM classification results: (e) Pixel-based, (f) SCRM, (g) MHIS, and (h) Proposed..... 139
7.35 (a) Slice #25 and 30 of an MRI volume "11_3", with corresponding (b) Gradient maps, (c) Volume growth maps, (d) Texture maps, and (e) final segmentation outcomes 142
7.36 Gray/white-matter segmentation results: (a) ITK-SNAP (b) Proposed algorithm, and (c) Ground Truth segmentations 144
7.37 Target sub-volume - Lungs: (a) Arbitrary CT slice (left) and user-defined line markings in the SVOI (right), (b) 2-D view of the corresponding segmentation using the proposed algorithm across two different slices, and (c) 3-D view of the segmentation results using the proposed method (left) and ITK-SNAP (right). 145
7.38 Target sub-volume - Spleen: (a) Arbitrary CT slice (left) and user-defined line markings in the SVOI (right), (b) 2-D view of the corresponding segmentation using the proposed algorithm across

two different slices, and (c) 3-D view of the segmentation results using the proposed method (left) and ITK-SNAP (right) 147
7.39 Target sub-volume – Vertebral structure: (a) Arbitrary CT slice (left) and user-defined line markings in the SVOI (right), (b) 2-D view of the corresponding segmentation using the proposed algorithm across two different slices, and (c) 3-D view of the segmentation results using the proposed method (left) and ITK-SNAP (right) 148
8.1 Fundamental principle of operation of the proposed FMV segmentation algorithm. 154
8.2 Segmentation results: (a) input video frames, (b) EHGVS algorithm, and (c) proposed algorithm..... 155
8.3 Segmentation results acquired from a synthetic FMV airborne sequence simulated using DIRSIG. 157

List of Tables

TABLE 7.1: EVALUATION OF SEGMENTATION RESULTS. 116
TABLE 7.2: NPR EVALUATION OF SEGMENTATION RESULTS USING 300 IMAGES OF THE BERKELEY DATABASE. 125
TABLE 7.3: INDIVIDUAL CLASSIFICATION ACCURACIES (%) FOR THE UNIVERSITY IMAGE 140
TABLE 7.4: INDIVIDUAL CLASSIFICATION ACCURACIES (%) FOR THE CENTER IMAGE 140
TABLE 7.5: GLOBAL CLASSIFICATION ACCURACIES (%) FOR THE UNIVERSITY IMAGE. 141
TABLE 7.6: GLOBAL CLASSIFICATION ACCURACIES (%) FOR THE CENTER IMAGE. 141
TABLE 7.7: GRAY- MATTER SEGMENTATION ACCURACIES (%) 149
TABLE 7.8: WHITE- MATTER SEGMENTATION ACCURACIES (%) 150

Chapter 1

Introduction

The interest in digital media has grown to new heights with rapid technological advancements being made in the capture and sharing of image data, a scenario that is applicable to most imaging modalities. This has necessitated the exploration of methods to interpret, enhance, classify and/or extract information from them. Segmentation is one approach that provides the foundation to make these functionalities ever more effective and expeditious by separating image data into their individual constituents.

1.1 Problem Statement and Motivations

Segmentation facilitates separation of 2-D / 3-D information contained in image data into their individual constituents, a task that is accomplished quite comfortably by our visual system and cortical mechanisms. However, mimicking this capability of human observers in an artificial or simulated environment has been found to be an extremely challenging problem. Formally, for the 2-D case, segmentation is defined as the process of partitioning or segregating picture elements (called pixels) of an image data set into non-overlapping regions (also called as clusters or groups) manifesting homogeneous or nearly homogeneous attributes such as intensity, texture, gradient as well as spatial attributes pertaining to location. On the other hand, for 3-D image data, segmentation is defined as the process of partitioning volumetric elements (called voxels) into non-overlapping sub-volumes exhibiting homogeneous or nearly homogeneous attributes. At a fundamental level a segmentation routine for a dataset is said to be "complete" when it provides an outcome comprising of a unique region/sub-volume label assignment for every data element, such that all pixels/voxels associated with a segmented region/sub-volume satisfy certain criteria while the same criterions are not universally satisfied for pixels/voxels from disjoint regions/sub-volumes. In this research the outcome of a segmentation routine will be provided as a labeled 2-D/3-D matrix of numbers each explicitly associated with a randomly generated pseudo color representing a specific segment, as seen in the images of Fig. 1.1.

The cardinal motivation for 2-D / 3-D image data segmentation is two-fold. It not only provides an end user with the flexibility to efficiently access and manipulate individual content, but also furnishes a fairly compact representation of the data where-in all subsequent processing can be done at a region/sub-volume level as opposed to the pixel/voxel level, resulting in large computational gains. To this effect, segmentation is predominantly employed as a pre-processing step to annotate, enhance, categorize and abstract information from images.

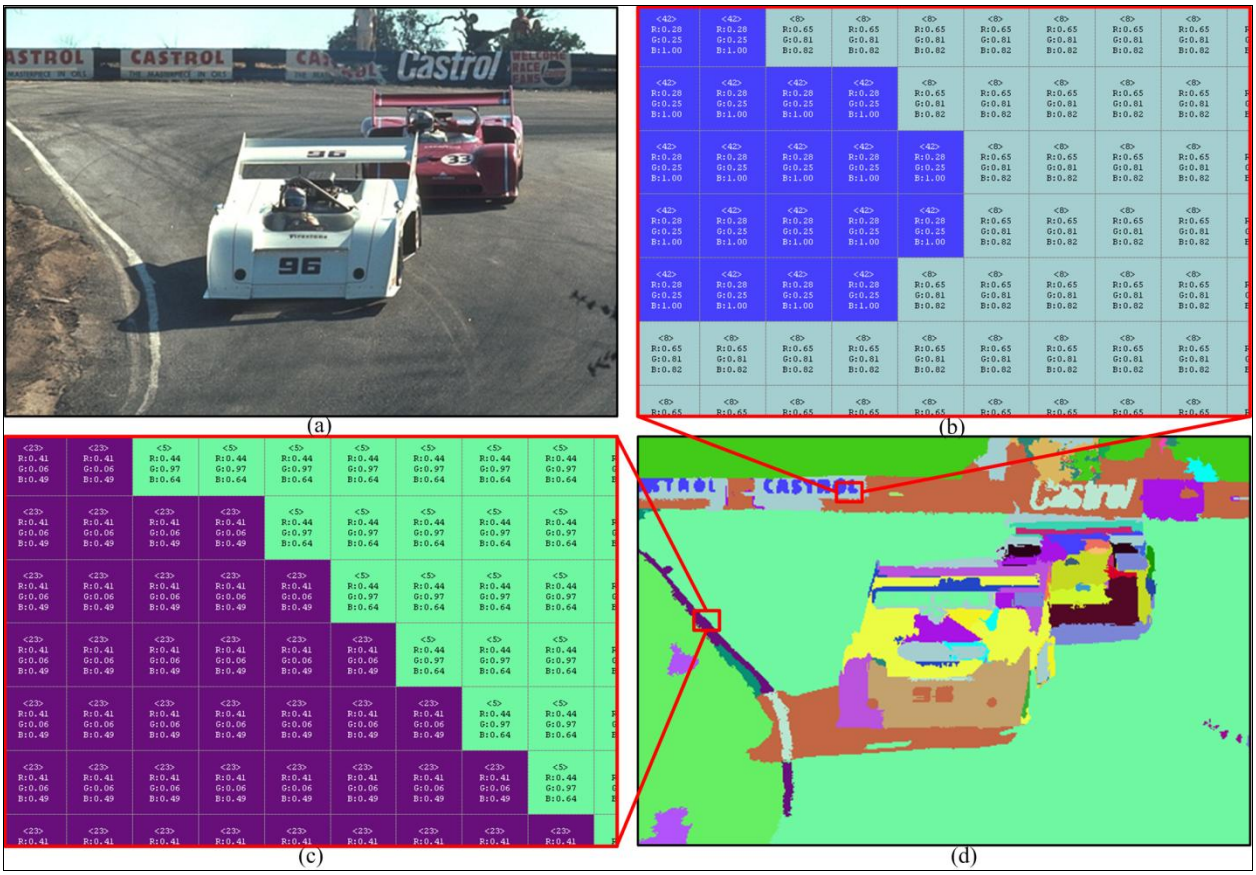

Fig. 1.1: (a) Color image, (b, c) show a matrix of region labels at two arbitrary locations in the image, (d) a complete pseudo colored segmentation result corresponding to the input color image.

In general, there are many application areas for segmentation pertinent to color imagery such as content-based image retrieval (CBIR), image rendering, region classification, segment based compression, surveillance, perceptual ranking of regions, graphics, and multimedia applications. Many approaches developed for color imagery have also found its use in other modalities of imaging such as remote sensing and biomedical imaging for highly sophisticated geo-spatial and medical intelligence applications pertinent to 3-D object modeling, large area search, change detection, and 3-D visualization/navigation to name a few. The exponential growth of the number of the applications for segmentation in itself provides a strong motivation for continued research and development.

Segmentation is often viewed as an ill-defined problem with no perfect solution but multiple generally acceptable solutions due to its subjective nature. The subjectivity of segmentation has been extensively substantiated in experiments [Martin et al., 2001] conducted in the color domain at the University of California at Berkeley to develop a benchmark for segmentation evaluation, where-in a database of manually generated segmentations using multiple human observers for a set of color images with natural content were obtained. In Fig. 1.2 (a), ten images (arbitrarily named airplane, starfish, race cars, hills, boat, church, cheetah, dolphins, lake and skydiver) from the aforementioned database are

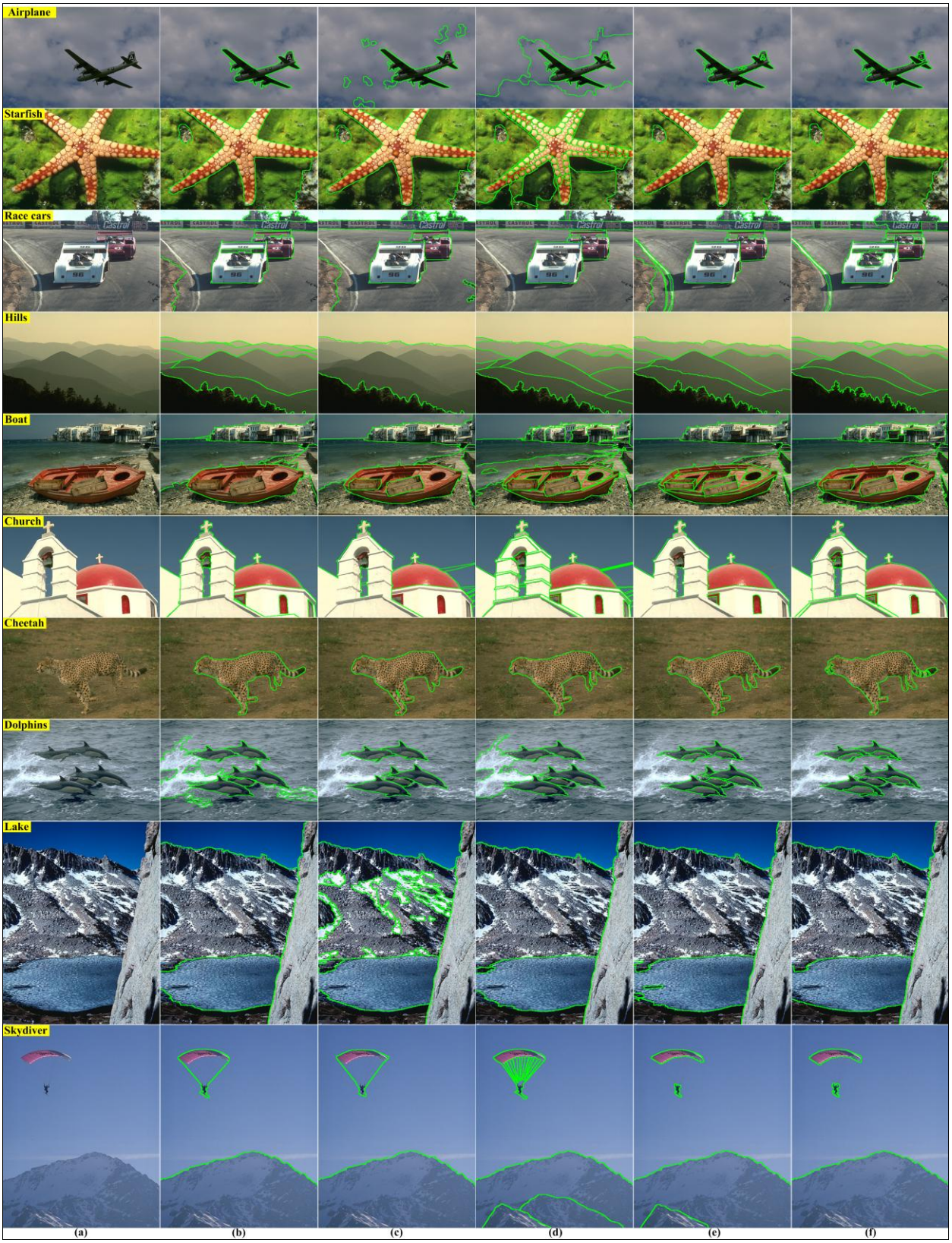

Fig. 1.2: Berkeley segmentation benchmark [Martin et al., 2001]: (a) Original images, and (b)-(f) region boundaries of multiple manually generated segmentations overlaid on the images.

displayed. Additionally, several manually segmented ground truths with region boundaries superimposed (in green) on the original image are shown in Fig. 1.2 (b)-(f). Analysis of the obtained ground truth results by Martin et al. divulged two imperative aspects: 1) an arbitrary image may have a unique acceptable segmentation outcome while there might be others that have multiple acceptable solutions, and 2) the variability in accepted solutions is primarily due to differences in the level of attention (or granularity) and degree of detail of perception of images from one human observer to another. Consequently, most present day algorithms for segmentation aim at providing generally acceptable outcomes rather than a "gold standard" solution. However it's important to note that the degree of subjectivity of segmentation depends on the imaging modality and the problem at hand. For example, if a certain application entails segmenting specific or well-defined content from a dataset, then the subjective nature of the outcomes could be significantly lower than applications that require segmentation of the entire dataset.

1.2 Research Goals

From a high level perspective (see Fig. 1.3), the primary objective of this doctoral study is to develop efficient algorithms (using a common framework) capable of segmenting 2-D / 3-D image data, acquired from different modalities. More specifically, three target imaging modalities pertaining to conventional color (3-band) or photographic imagery, remotely sensed multi/hyperspectral imagery (> 3 bands) and bio-medical volumes pertinent to Magnetic Resonance Imaging (MRI) and Computed Tomography (CT), have been used in this work, as seen in Fig. 1.3. Listed below is a detailed description of the complete set of objectives that have been accomplished this doctoral study:

1. A novel unsupervised color image segmentation algorithm called GSEG has been developed, which exploits the information obtained from detecting edges in color images in a single scale framework. This Gradient-based SEGmentation (GSEG) method is initialized by a vector gradient calculation in the CIE L*a*b* color space. The obtained gradient map is utilized for initially clustering low gradient content, as well as automatically generating thresholds for a computationally efficient dynamic region growth procedure, to segment regions of subsequent higher gradient densities in the image. The resultant segmentation is combined with an entropy-based texture model in a statistical merging procedure to obtain the final result.
2. A novel unsupervised multiresolution color image segmentation algorithm called MAPGSEG (Multi-resolution Adaptive and Progressive Gradient based color image SEGmentation) has been instituted, which takes advantage of gradient information in an adaptive and progressive framework. The proposed methodology is initiated with a dyadic wavelet decomposition scheme of an arbitrary input image, accompanied by a vector gradient calculation of its color converted counterpart. The resultant gradient map is used to automatically and adaptively generate

thresholds for segregating regions of varying gradient densities, at different resolution levels of the input image pyramid. At each level, the classification obtained by a progressively thresholded growth procedure is integrated with an entropy-based texture model utilizing a unique region merging procedure to obtain an interim segmentation. In combination with a confidence map and non-linear spatial filtering techniques, regions of high confidence are passed from one resolution level to another until the final segmentation at highest (original) resolution is achieved.

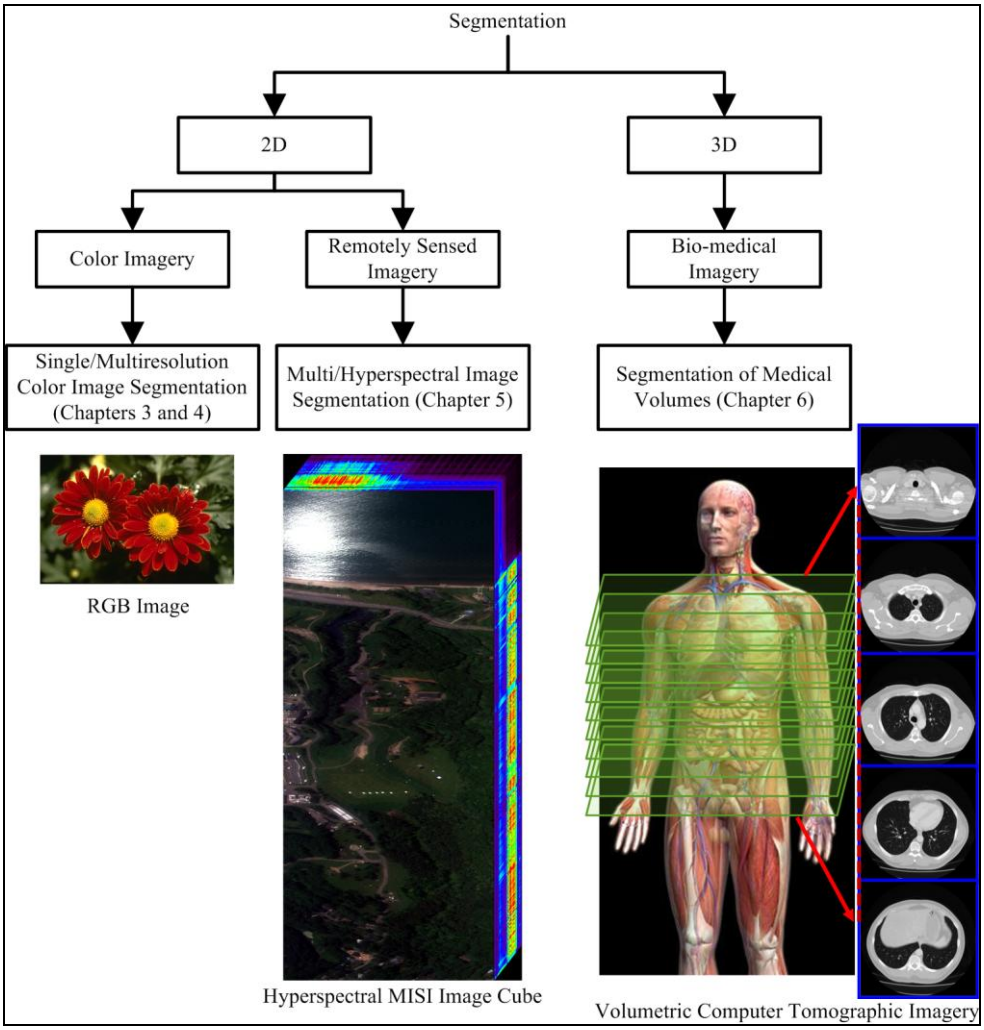

Fig. 1.3: A bird's eye view of research goals

3. A novel unsupervised algorithm that efficiently utilizes information derived from spectral, gradient and textural attributes for spatially segmenting multi/hyperspectral remotely sensed imagery has been proposed. Our methodology commences by determining the magnitude of spectral intensity variations across the input scene, using a multiband gradient detection scheme optimized for handling remotely sensed image data. The resultant gradient map is employed in a

dynamic region growth process that is initiated in pixel locations with small gradient magnitudes and is concluded at sites with large gradient magnitudes, yielding a map comprising of an initial set of regions. This region map is combined with several co-occurrence matrix-derived textural descriptors along with intensity and gradient features, in a multivariate analysis-based region merging procedure that fuses regions with similar characteristics to yield the final segmentation output. Data employed for this accomplishing research objective has been acquired from a wide range of multispectral/hyperspectral sensors such as the HYperspectral Digital Imagery Collection Experiment (HYDICE), Hyperion, Modular Imaging Spectrometer Instrument (MISI), and the Reflective Optics System Imaging Spectrometer (ROSIS) etc., to name a few.

4. A novel technique that efficiently employs intensity, gradient and textural features for three-dimensional (3-D) segmentation of medical (MRI/CT) volumes, has been introduced. Our methodology commences by determining the magnitude of intensity variations across the input volume, using a 3-D gradient detection scheme. The resultant gradient volume is employed in a dynamic volume growing/formation process that is initiated in voxel locations with small gradient magnitudes and is concluded at sites with large gradient magnitudes, yielding a map comprising of an initial set of partitions (or sub-volumes). This partition map is combined with an entropy based texture descriptor along with intensity and gradient attributes, in a multivariate analysis-based volume merging procedure that fuses sub-volumes with similar characteristics to yield a final/refined segmentation output. Additionally, this work also involved the development of a semi-automated version of the aforestated algorithm that allows a user to interactively segment a desired Sub-Volume of Interest (SVOI) as opposed to the entire volume. The data being employed for this research has been provided by the Massachusetts General Hospital, National Biomedical Imaging Archive (NBIA), and DataPhysics Research (DPR).

1.3 Original Contributions

The primary original contribution of this research (from a holistic viewpoint) is that it facilitates a single effective framework for segmenting 2-D/3-D image data acquired from different imaging modalities. Other important contributions made to the segmentation realm are as follows:

1. Most segmentation algorithms are generally based on one of two basic properties of image data: a) discontinuity, and b) similarity. The core segmentation framework proposed in this research effort has been employed in three target imaging modalities and is unique in the fact that it simultaneously lays emphasis on the homogenous and heterogeneous characteristics of image data using multiple attributes.

2. A unique region/volume growing approach has been designed based on the principle of dynamic gradient thresholding that iteratively thresholds gradient information, commencing at pixel/voxel locations with small gradient magnitudes (gradually varying intensity/radiodensity) and culminating at pixels/voxel locations of large gradient magnitudes (abrupt intensity/radiodensity variations) with no (a) dependency on the order in which they are processed, (b) manual selection of seed points, and/or (c) pixel-by-pixel agglomeration. Another aspect of our growth approach that is different from conventional approaches is its parallel processing capabilities of simultaneously growing multiple adjacent/non-adjacent image regions / sub-volumes.
3. An efficient approach for the merging of regions / sub-volumes based on statistical analysis of a multi-attribute space has been introduced, facilitating a mechanism to fuse information from distinct features to define meaningful segments, even in the presence of texture disparities, spatial separation and occlusion.
4. The segmentation framework established in this work does not exclusively depend on the initial assignment of clusters, to arrive at the final segmentation. In fact, if need be, the approach has the capability of producing a suite of segmentation outcomes ranging from over-segmented ones to under-segmented results.
5. A multi-resolution segmentation algorithm (MAPGSEG) has been developed, that facilitates: a) selective access and manipulation of individual content in images based on desired level of detail, b) treatment of sub sampled versions of images with robustness to scalability, c) a potential solution that computationally measures up to meet the demands of most practical applications involving segmentation, and d) a practical compromise between quality and speed, laying the foundation for fast and intelligent object/region based real-world applications of color imagery. Through the aforestated technique it is demonstrated that the segmentation results of low-resolution images can be utilized to efficiently segment their corresponding high resolution counterparts.
6. An efficient approach for achieving improved localization of edges detected in remotely sensed imagery has been introduced, wherein the improvement is in the localization of the detected edges. Furthermore, a mathematical framework for gradient detection in 3-D data is provided. Although the previously mentioned vector gradient approach has been used extensively in the 2-D domain, this work is the first attempt to provide and demonstrate a 3-D extension.
7. Preliminary demonstration of the proposed algorithm for handling full motion digital video (see Section 8.3.1).
8. Finally, several segmentation-related literary contributions have been made (see Section 1.5).

1.4 Intellectual Merit and Broader Impacts

Content manipulation in image data has been growing at an unprecedented rate over the last decade. Lately, image content is being uploaded and viewed at a torrential rate via user-friendly websites such as facebook, twitter, and Google+ to name a few. In fact every minute, large volumes of imagery are uploaded on the web contributing to a multi-billion dollar industry. In general, these databases of imagery involving 2-D/3-D content contain a variety of intensity, texture, and gradient information that require critical and custom handling to achieve optimum visual/application/system performance. Most often, consumers are interested in quickly and effectively "manipulating" portions of the content to achieve a given objective. For instance, one may be interested in adjusting the flesh-tone or memory colors in conventional color imagery or creating a skeleton look of a given volume through edge detection in 3-D. Applications like these are growing at a swift pace in our current environment given the rapid advances being made in digital technologies. At present, most consumer-based multimedia systems do not possess the capabilities to manipulate image content in 2-D or 3-D in a flexible and effective manner without the need for sophisticated and expensive software packages and constrained settings. The bounty of data has rendered the development of good automated/semi-automated multimedia communication systems/tools to have fallen far behind.

Consequently, intellectual benefits of the proposed research are multi-fold. It will provide the end user with the ability to selectively access and manipulate 2-D/3-D image data obtained from disparate modalities by providing meaningful segmentations in the presence of contrasting intensity variations, texture disparities and varying degrees of occlusion in a suitable time frame (real time or near real time), taking full advantage of multiple attributes derived from the scene/volume of interest. If successful the proposed work can be used for widespread commercial and research applications in the fields of computer vision, image processing, graphics, animation, multimedia communications, human-computer interactions, security/surveillance, tracking and any application requiring rapid automated low and mid level analysis. The proposed work has a broader impact along the following aspects: 1) potential financial gain from Intellectual Property protection (patents), and 2) increase of market awareness by publishing our results and findings. In addition the outcomes of this research will be made available to stakeholders involving the Rochester Institute of Technology and its collaborators in the form of reports, MATLAB software for internal/ confidential use, and presentations as appropriate. At the request of stakeholders, software developed will be made available for future teaching as well as research endeavors.

1.5 Publications and Intellectual Property representing Original Contributions

Listed below are the set of peer reviewed journal publications that signify the original contributions made in course of this dissertation:

1. S. R. Vantaram, E. Saber, S. A. Dianat, and Y. Hu, "Synthesis of Intensity Gradient and Texture Information for Efficient Three-Dimensional Segmentation of Medical Volumes", *under review, IEEE Transactions on Medical Imaging.*
2. S. R. Vantaram, E. Saber, and D. Messinger, "Automatic Spatial Segmentation of Multi/Hyperspectral Imagery by Fusion of Spectral-Gradient-Textural Attributes", *under review, IEEE Transactions on Geoscience and Remote Sensing*.
3. S. R. Vantaram, and E. Saber, "A Survey of Contemporary Trends in Color Image Segmentation", *Journal Electronic Imaging, Volume 21, Number 4, October-December 2012.*
4. S. R. Vantaram, E.Saber, S. A. Dianat, M. Shaw, R. Bhaskar, "Multiresolution Adaptive and Progressive Gradient-based color image SEGmentation", *Journal of Electronic Imaging, Volume 19, Number 1, pp. 013001 (1-21), January-March 2010.*
5. L. E. Garcia, E. Saber, S. R. Vantaram, V. Amuso, M. Shaw, R. Bhaskar, "Automatic Image Segmentation by Dynamic Region Growth and Multiresolution Merging", *IEEE Transactions on Image Processing, Volume 18, Number 10, pp. 2275-2288, October 2009.*

Listed below are the set of conference publications that signify the original contributions made in course of this dissertation:

1. S. R. Vantaram, and E. Saber, "Unsupervised Video Segmentation by Dynamic Volume Growing and Multivariate Volume Merging Using Color-Texture-Gradient Features", *accepted, 19th IEEE International Conference on Image Processing, Orlando, FL, September 2012 (Refereed).*
2. S. R. Vantaram, E. Saber, S. A. Dianat, Y. Hu and V. Abhyankar, "Semi-Automatic 3-D Segmentation of Computed Tomographic Imagery by Iterative Gradient-Driven Volume Growing", *18th IEEE International Conference on Image Processing, Brussels, Belgium, September 2011 (Refereed).*
3. S. R. Vantaram, E. Saber and D. Messinger, "An Intensity-Gradient-Texture Guided Methodology for Automatic Spatial Segmentation of Remotely Sensed Multi / Hyperspectral imagery", *18th IEEE International Conference on Image Processing, Brussels, Belgium, September 2011 (Refereed).*
4. S. R. Vantaram, and E.Saber, "An Approach for Improved Localization of Edges in Multi-Band Imagery", *Proceedings of the SPIE Optical Engineering +Applications Conference, San Diego, CA, Aug. 2011.*
5. S. R. Vantaram, and E.Saber, "An Adaptive Bayesian Clustering and Multivariate Region Merging-based Technique for Efficient Segmentation of Color Images", *36th IEEE International*

Conference on Acoustics, Speech and Signal Processing, Prague, Czech Republic, May 2011 (Refereed).

6. M. Jaber, S. R. Vantaram, and E.Saber, "A Probabilistic Framework for Unsupervised Evaluation and Ranking of Image Segmentations", *39th IEEE Applied Imagery Pattern Recognition Workshop, Washington DC, October 2010.*
7. S. R. Vantaram, E.Saber, S. A. Dianat, M. Shaw, and R. Bhaskar, "An Adaptive and Progressive approach for Efficient Gradient-Based Multiresolution Color Image Segmentation", *16th IEEE International Conference on Image Processing, Cairo, Egypt, November 2009 (Refereed).*
8. S. R. Vantaram, E.Saber, V. Amuso, M. Shaw, and R. Bhaskar, "Unsupervised Image Segmentation by Automatic Gradient Thresholding for Dynamic Region Growth in the CIE L*a*b* Color Space", *Proceedings of 21st Annual SPIE/IS&T: Human Vision and Electronic Imaging Symposium, Volume 7240, pp. 724019 (1-11), San Jose, CA, January 2009.*

Listed below are the set of filed United States Patent applications that signify the original contributions made in course of this dissertation:

1. Vishwas Abhyankar, S. R Vantaram, E.Saber, Sohail Dianat, and Yang Hu, "Semi-Automatic 3-D Segmentation of Computed Tomographic Imagery by Iterative Gradient-Driven Volume Growing", U.S. Provisional Patent Application Attorney Docket # DTPHPZ00300 (filed by the DataPhysics Research Incorporation).
2. S. R Vantaram, E. Saber, S. A. Dianat, M. Shaw, and R. Bhaskar, "Methods for adaptive and progressive gradient-based multi-resolution color image segmentation and systems thereof", U.S. Patent Application Serial No. 12/655,933 (jointly filed by the Rochester Institute of Technology and the Hewlett Packard Company).

1.6 Organization

The remainder of this document is organized as follows (see Fig. 1.3). Chapter 2 provides a general/broad categorization of segmentation approaches employed in the color domain as well as specific categorization based on the methodology used. This chapter also provides a literature review of methods adopted for segmentation of multi/hyperspectral imagery and MRI/CT data in 3-D. The GSEG algorithm which forms our core segmentation framework is described in Chapter 3. The MAPGSEG algorithm developed to perform color image segmentation in a multiresolution framework is described in Chapter 4. In Chapter 5, a methodology for segmentation of multi/hyperspectral imagery is detailed, while a description of a segmentation algorithm for 3-D medical volumes is provided in Chapter 6. Experimental

results obtained so far are provided in Chapter 7. Finally, a summary of research accomplishments, conclusions and future directives are discussed in Chapter 8.

Chapter 2

Literature Review

In this section, an overview of various 2-D/3-D methodologies developed to accomplish the segmentation objective for conventional photographic images, remotely sensed data and medical volumes, is provided. Segmentation procedures can be broadly categorized from a high level perspective as well as specifically grouped based on their technical grounding (low level classification). The following sub-sections describe each of the two taxonomies applicable to the aforementioned modalities of imaging, in detail.

2.1. High Level Taxonomy

In general segmentation techniques can be broadly classified (see Fig. 2.1) based on the: 1) image type, 2) extent of human interaction, 3) manner in which the image is represented for processing, 4) number and type of attributes used, and 5) fundamental principle of operation.

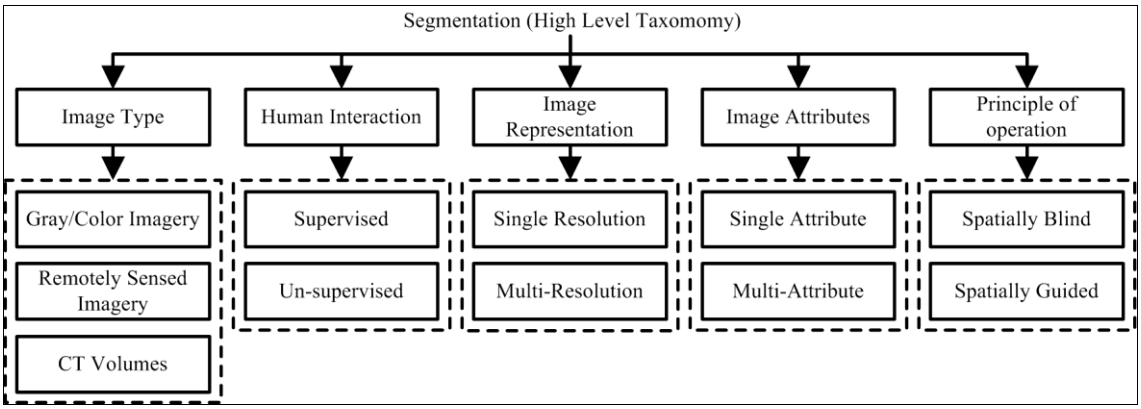

Fig. 2.1: High level taxonomy of image segmentation algorithms.

The first criterion segregates algorithms that have been developed for data acquired from a specific imaging modality. In this work, the "image type" criterion segregates algorithms developed for monochrome/color images from the ones that handle multi/hyperspectral data as well as 3-D medical volumes such as MRI/CT data. The second criterion distinguishes algorithms that require human intervention (supervised processes) for segmentation from the ones that operate without any manual interference (fully automatic or unsupervised processes). The third criterion separates segmentation procedures that directly operate on the original image/volume (single resolution configuration) from the ones that operate on multiple representations of the image/volume (multiresolution configuration), each yielding different amount of the information. The fourth criterion differentiates algorithms based on the type of image/volume information (such as intensity, radiodensity, gradient, or textural features, to name

a few) they utilize to perform segmentation. It is imperative to note that most methods use the aforementioned attributes individually (single attribute methods) or in specific combinations (multiple attribute methods) which can be used to categorize them. Finally, the last criterion based on the underlying principle of operation discriminates segmentation mechanisms as being either spatially blind or spatially guided. Spatially blind approaches as the name suggests are techniques that are "blind" to spatial information or in other words do not take into account the spatial arrangement of pixels/voxels in an image/volume. Instead these methods heavily rely on grouping image information in suitable attribute/ feature spaces. On the other hand spatially guided approaches take into consideration the spatial arrangement of pixels/voxels in the data during the segmentation process.

2.2. Low Level Taxonomy

This sub-section discusses segmentation procedures based on their technical components separately for the three imaging modalities pertinent to this doctoral research.

2.2.1 Color/Gray Scale Image Segmentation Methodologies

There are several excellent surveys of color/gray scale image segmentation strategies and practices. The studies done by Fu et al. [Fu et al., 1981] and Pal et al. [Pal et al., 1993] are amongst the earliest ones that have been widely popular. In their work, Fu et al. categorized segmentation approaches developed during the 70's and early 80's for gray scale images into three classes namely, clustering, edge detection, and region extraction. On the other hand, Pal et al. reviewed more complex segmentation techniques established amid the late 80's and early 90's that involved fuzzy/non-fuzzy mechanisms, Markov Random Fields (MRFs) probabilistic models, color information as well as neural networks; all of which were still in their early stages of development. The surveys done by Lucchese et al. [Lucchese et al., 2001] and Cheng et al. [Cheng et al., 2001] were amongst the first that exclusively provided an in-depth overview of algorithms targeted at segmenting color images, instituted throughout the 90's.

As mentioned previously, most segmentation modus operandi can be viewed as being either spatially blind or spatially guided. It is this distinction that forms the basis of our low-level taxonomy where we specifically group color/gray scale image segmentation procedures based on their technical components, as depicted in Fig. 2.2.

2.2.1.1 Spatially Blind Approaches

Spatially blind approaches perform segmentation in certain attribute/feature spaces, predominantly related to intensity (gray or color). Popular segmentation techniques that fall within the notion of being spatially blind involve clustering and histogram thresholding.

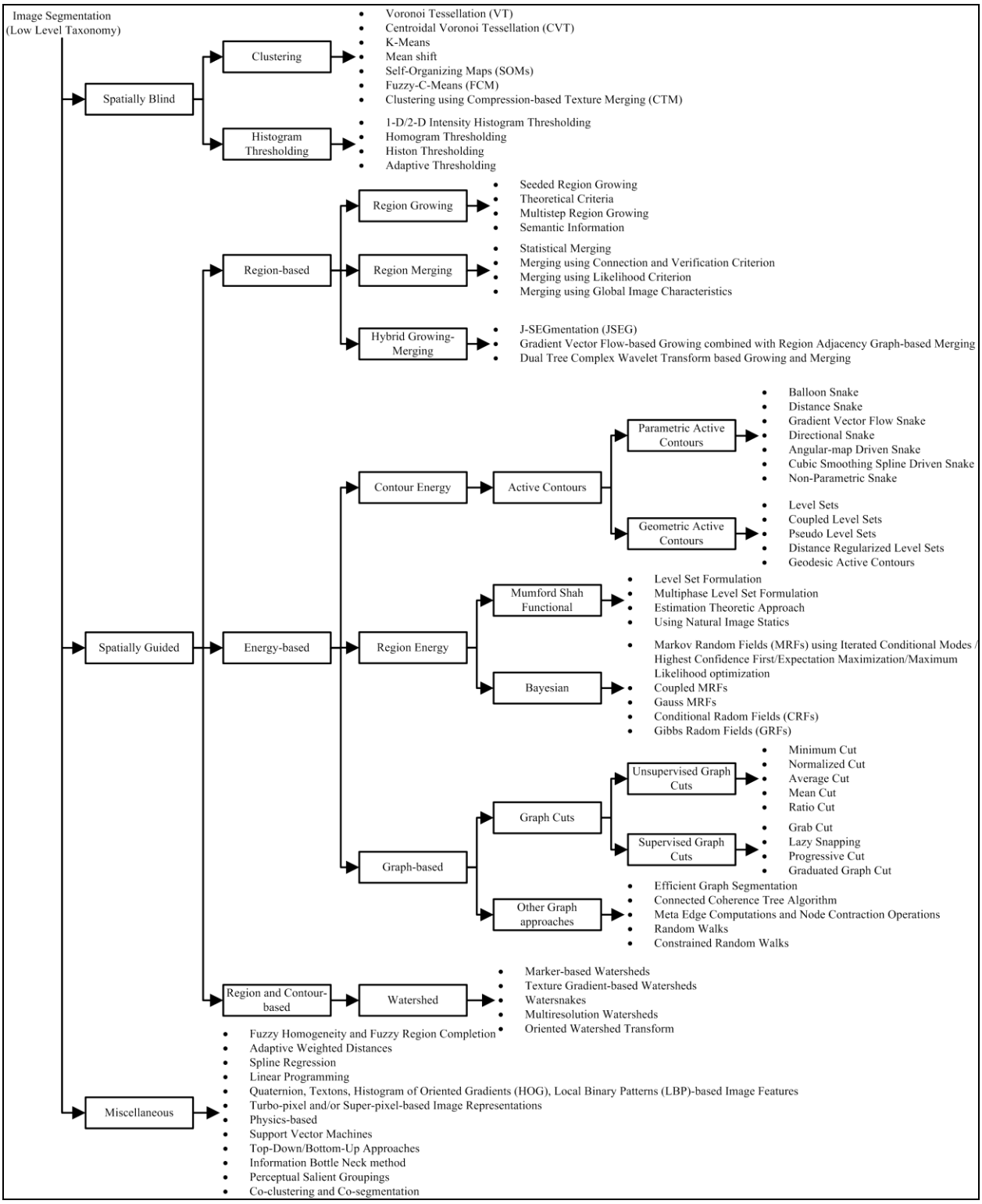

Fig. 2.2: Low level taxonomy of color/gray scale image segmentation algorithms.

2.2.1.1.1 Clustering

In its simplest form, clustering is a spatially blind technique wherein the image data is viewed as a point cloud on a 1-D gray scale axis or in a 3-D color space (see Fig. 2.3) depending on the image type.

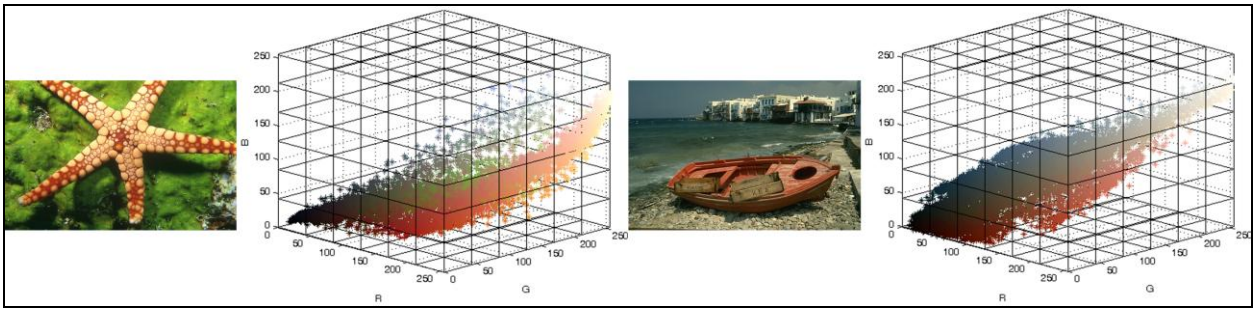

Fig. 2.3: Sample color images with their corresponding 3-D point clouds.

Several different color spaces - such as RGB, Commission International de l'Eclairage (CIE) L*a*b* and L*u*v*, YC_bC_r, HSI etc., to name a few - with different properties have been extensively utilized for segmentation [Green et al., 2002]. The essence of a typical clustering protocol is to analyze this gray/color intensity point cloud and partition it using pre-defined metrics/objective functions to identify meaningful pixel groupings also known as classes or clusters. Furthermore, the clustering process is done such that, when complete, the pixel data within a specific class possess, in general, a high degree of similarity while the data between classes has low affinity. The biggest advantage of clustering approaches over others is inherent in their simplicity and ease of implementation. However, since the point clouds generated are image dependent, selecting/initializing the number of clusters so as to obtain semantic partitioning with respect to the image being processed is a challenging task, especially in the case of color imagery. Furthermore, as the dimensionality of the feature space is increased exponentially, acquiring definitive clusters becomes an arduous task.

Although many clustering approaches have been developed for various applications, partitioning a feature space using a specific set of fixed points is the most widely adopted approach. Voronoi tessellation (VT) is a procedure in which a feature space is decomposed into various clusters (called Voronoi cells/regions) using a fixed set of points called sites, such that each observation in the feature space is assigned to the closest site predicated on a certain distance metric. More specifically, if X is a feature space constrained with a distance function d, and $(P_k)_{k\in K}$ is a set of K sites in the space, then a Voronoi cell V_k formed using the site P_k is the set of all points $x \in X$ that satisfy:

$$V_k = \{x \in X \mid d(x, P_k) \leq d(x, P_j)\ \forall j \neq k\} \tag{2.1}$$

where $d(x, P_k)$ represents the distance from x to P_k. Arbeláez et al. [Arbeláez et al., 2006] proposed a VT-based image segmentation technique utilizing color and lightness information derived from the image. The segmentation objective was achieved in a two-step process comprised of: 1) pre-segmentation

and 2) hierarchical representation. The pre-segmentation step employed a VT process wherein the extreme components of the lightness (L*) channel were used as sites, to form an extrema mosaic of Voronoi regions. The second step involved the development of a stratified hierarchy of partitions derived from the extrema mosaic using a pseudo-distance metric called ultrametric, specifically defined for color images. Subsequently, a single real valued soft boundary image called the Ultra-metric Contour Map (UCM) was constructed to arrive at the final segmentation.

Centroidal Voronoi Tessellation (CVT) is a special category of VT wherein the sites producing Voronoi cells are chosen equivalent to their center of mass. Wang et al. [Wang et al., 2009] generalized the basic CVT by integrating an edge related energy function with a classic clustering energy metric, to propose an Edge Weighted Centroidal Voronoi Tessellation (EWCVT) for effective segmentation of color images. CVTs form the core of many prominent clustering algorithms such as K-means. The K-means algorithm partitions a set of n-pixels into K clusters by minimizing an objective function. From a color segmentation perspective, the aforementioned algorithm analyzes the image data in a 3-D space to eventually identify K-sites (known as cluster centers or means) such that the mean squared distance from each data point to its nearest center is minimized. To this effect, in an arbitrary iteration (called as a Voronoi iteration or Lloyd's algorithm), the entire color space is separated into K partitions by assigning each observation to the cluster with the closest center (Note initialization in the first iteration may be randomly done). Following this, a new estimate of the cluster center is computed based on the current cluster assignment information, and utilized as an input to the next iteration of the algorithm. The algorithmic steps described above are repeated until convergence is achieved. McQueen [McQueen, 1967] was the first to employ the K-means algorithm to handle multivariate data. Among recent advances, Kanungo et al. [Kanungo et al., 2002] proposed an efficient version of the algorithm called the "filtering algorithm", by utilizing a k-dimensional (kd) tree representation of the image data. For each node of this tree, a set of candidate centers were determined and strategically filtered as they were propagated to the node's children. The biggest advantage of this approach was that, since the kd-tree representation was formed from the original data rather than from the computed centers, it did not mandate an update in its structure for all iterations, in contrast to the conventional K-means architecture. Chen et al. [Chen et al., 2005] employed a generalization of the classical K-means algorithm better known as the Adaptive Clustering Algorithm (ACA), with spatially adaptive features pertaining to color and texture, to yield perceptually tuned segmentations. Consequently, the ACA method is an exception to the norm of spatially blind clustering. In his work, Mignotte [Mignotte, 2008] proposed a novel color image segmentation procedure based on the fusion of multiple K-means clustering results by minimizing the Euclidean distance function, obtained from an image described in six different color spaces namely RGB, HSV, YIQ, XYZ, LAB, and LUV. Once the label fields from each of these color spaces are

obtained, a local histogram of the class labels across the aforementioned label fields is computed for each pixel and the set of all histograms are employed as input feature vectors to a fusion mechanism that culminates in the final segmentation output. The fusion scheme is comprised of the K-means algorithm using the Bhattacharya similarity coefficient, which is a histogram-based similarity metric. The algorithm in [Mignotte, 2008] was further enhanced in [Mignotte, 2011] by using a spatially constrained K-means labeling process in place of the fusion mechanism to arrive at the optimal result. While the prior algorithm developed by Mignotte was aimed at exploring the possibility of integrating multiple segmentation maps from simple data partitioning models to obtain an accurate result, the later algorithm was novel in the sense that within the K-means framework implicit spatial associations in an image were taken into account (similar to the work in [Chen et al., 2005]) to uncover the best solution, and consequently the process was not completely spatially blind.

Mean shift clustering [Comaniciu et al., 2002] is another routine that has had pervasive use for gray/color image segmentation within the last decade. This generic non-parametric technique facilitates the analysis of multidimensional feature spaces with arbitrarily shaped clusters, based on the "mean shift" concept, originally proposed by Fununaga et al. [Fununaga et al., 1975]. The mean shift procedure is a kernel density estimation (or Parzen window)-based technique that scrutinizes a feature space as an empirical probability density function (pdf), and considers the set of pixel values from an arbitrary image as discrete samples of that function. The procedure exploits the fact that clusters/dense regions in a feature space typically manifest themselves as modes of the aforestated pdf. In what follows, if S is a finite point cloud in an n-dimensional Euclidean space, X and K is a symmetric Kernel function of specific characteristics, then the sample mean $m(x)$ at a pixel $x \in X$ computed utilizing a weighted combination of its nearby points determined by K is given as [Cheng, 1995]:

$$m(x) = \frac{\sum_{s \in S} K(s - x)s}{\sum_{s \in S} K(s - x)} \tag{2.2}$$

To this effect, at every pixel location x, a mean shift vector $m(x) - x$ is obtained with K centered at x, such that the vector points towards the direction of the maximum increase in density. Subsequently, the operation $x \leftarrow m(x)$ is performed that shifts the value of x towards the mean followed by the re-estimation of $m(x)$. This process is repeated until convergence of $m(x)$ is achieved. At the end of the mean shift process, the closest peak in the pdf is identified for each pixel. Since the Mean shift algorithm uses spatial knowledge in its framework it also represents an exception to conventional spatially blind clustering. Mean shift clustering guided by edge information was first seen in the work by Christoudias et al. [Christoudias et al., 2002], who proposed the Edge Detection and Image Segmentation (EDISON)

system, aimed at improving the sensitivity of extracting homogeneous regions while maintaining or ideally minimizing over-segmentation of an image. Fig. 2.4 illustrates a few results of the EDISON system using default parameters (spatial band h_s =7, color band width h_r =6.5, and minimum region size M =20). Hong et al. [Hong et al., 2007] proposed an improved version of the mean shift segmentation algorithm by incorporating: 1) an enhanced technique for mode detection, 2) an optimized process for the global analysis of the locally identified modes, and 3) the elimination of textured areas in order to achieve stable results in various background conditions. Ozden et al. [Ozden et al., 2007] pioneered an effective technique that combined low–level color, spatial and texture features in the mean shift framework for color image segmentation.

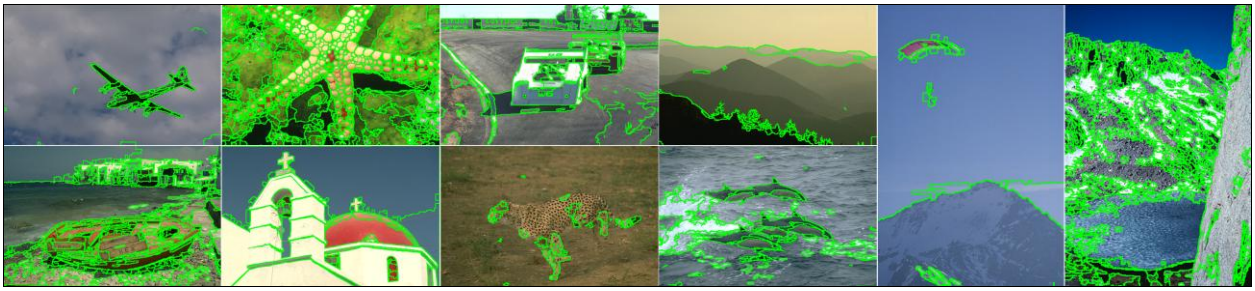

Fig. 2.4: Results of the EDISON system [Christoudias et al., 2002].

Neural networks-based data clustering is a category that has originated exclusively from the field of artificial intelligence. Within this domain, methods involving self-organizing maps (SOMs) have received the most attention in the last decade. A self-organizing map or a self-organizing feature map (SOFM) - alternately known as a Kohonen map - is a specific kind of artificial neural network (ANN) that was first introduced by Kohonen [Kohonen, 1990] as a tool for providing intelligent representations of high/multi-dimensional feature spaces in significantly lower (one or two) dimensions. A SOM (shown in Fig. 2.5) comprises of an input layer of nodes/neurons organized in a vector whose size is equivalent to the dimensions of the input feature space. Each node is connected in parallel to a 2-D output layer of neurons in a rectangular or hexagonal arrangement as well as their corresponding neighboring neurons utilizing an appropriate weighting scheme that signifies the strength of various connections. A SOM operates in a 'training' phase that gradually constructs a feature map using a sub-set of samples from the input feature space, followed by a 'mapping' routine in which an arbitrary new input sample is automatically classified. At the culmination of the two modes of operation, a low dimensional map that reflects the topological relationships of samples in the feature space predicated on their similarity is subsequently generated. In other words, samples that have similar characteristics in the input feature space form distinct clusters in this map.

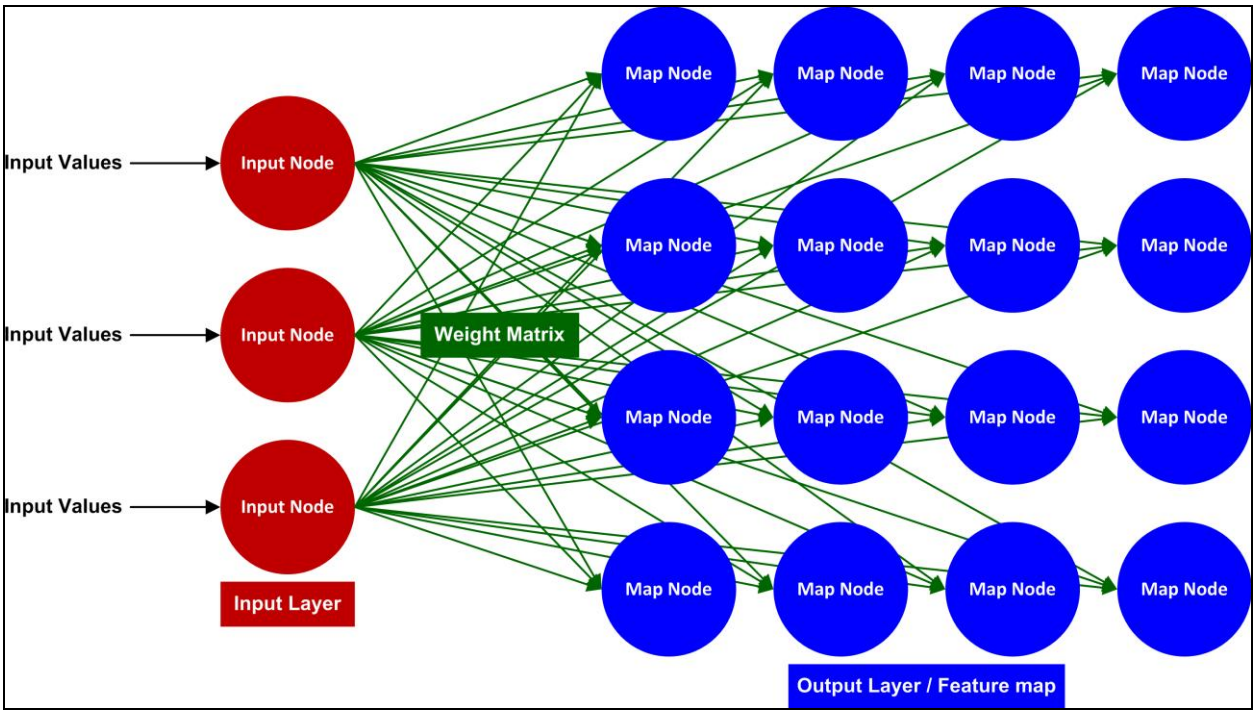

Fig. 2.5: Self-Organizing Map (SOM) in a rectangular neural layout.

Huang et al. [Huang et al., 2002] developed a color image segmentation methodology that employed a two stage SOM-based ANN. The algorithm is initiated by an RGB to HVC (Hue-Value-Chroma) color conversion of the input image which is employed by an SOM to identify a large initial set of color classes. The resultant set of classes are further refined by first computing the normalized Euclidean distance among them, and the obtained between-class distances are furnished as inputs to a second SOM that identifies the final batch of segmented clusters. In a similar scheme, Ong et al. [Ong et al., 2002] constructed a color image segmentation technique based on a hierarchical two stage SOM in which the first stage identifies dominant colors in the input image presented in the L*u*v* color space, while the second stage integrates a variable-sized 1-D feature map and cluster merging/discarding operations to acquire the eventual segmentation result. Li et al. [Li et al., 2003] demonstrated an effective color image segmentation approach using a SOM and the Fuzzy-C-Means (FCM) clustering procedure. The algorithm is initiated by finding well-suited image-dependent features derived from five different color spaces using a SOM. Subsequently, the obtained features were employed in a FCM protocol to attain the output segmentation. Dong et al. [Dong et al., 2005] instituted two alternate ANN-based strategies for color image segmentation. The first strategy was unsupervised. It involved distinguishing a set of color prototypes using SOM-based learning from the input image represented in the L*u*v* color space. These color prototypes were passed on to a simulated annealing-driven clustering routine to yield well defined clusters. The second method, built off the aforestated algorithm, was coupled with hierarchical pixel learning (that generated different sizes of color prototypes in the scene) and classification protocols to

provide more accurate segmentation outcomes in a supervised fashion. Partitioning of color imagery using SOM and adaptive resonance theory (ART) was first seen in the work of Yeo et al. [Yeo et al., 2005], who proposed two compound ANN architectures called SOMART and SmART (SOM unified with a variation of ART) that yielded improved segmentations in comparison to traditional SOM-based techniques. On the other hand, Araújo et al. [Araújo et al., 2009] designed a fast and robust self-adaptive topology ANN model called local adaptive receptive field SOM (LARFSOM) that deduced compact clusters and inferred their appropriate number based on color distributions learned rapidly from the network's training phase using a small percentage of pixels from the input image. The algorithm was tested on color images with varying segmentation complexities and was found to outperform several prior SOM-based techniques. Frisch [Frisch, 2006] introduced a novel approach robust to illumination variations that employed SOMs for the construction of fuzzy measures applicable to color image segmentation. This work was a unique attempt wherein efficient fuzzy measures, to accomplish the segmentation task, were derived using SOM-based processing. Ilea et al. [Ilea et al., 2008] devised a fully automatic image segmentation algorithm called CTex using color and texture descriptors. The CTex segmentation architecture first extracts dominant colors from the input image presented in the RGB and YIQ color spaces using an SOM classifier. In doing so, the appropriate number of clusters (K) in the scene are also identified. Subsequently, a conventional K -means clustering algorithm is employed in a 6-D multispace spanned by both the above stated color spaces, to obtain a segmentation result purely based on color information. This is followed by the computation of textural features using a Gabor filter bank, which along with the previously acquired segments, are provided as input to a novel adaptive spatial K -Means (ASKM) clustering algorithm that delineates coherent regions of color and texture in the input image.

The clustering techniques discussed so far are typically categorized as hard clustering approaches since every observation in the feature space has a unique and mandatory cluster assignment yielding clusters with sharp boundaries. In contrast, significant work has been done for the advancement of fuzzy clustering methods that facilitate observations to bear a certain degree of belongingness or membership to more than one cluster, resulting in overlapping clusters and/or clusters with "soft" boundaries. The Fuzzy-C-Means (FCM) algorithm, originally developed by [Dunn, 1974], is the most widely utilized fuzzy clustering methodology, and similar to the K -means technique, partitions a set of n -pixels ($X = \{x_1, \ldots, x_n\}$) into C fuzzy clusters ($C = \{c_1, \ldots, c_n\}$) by minimizing an objective function. The objective function utilized by the FCM algorithm is represented as:

$$J_m = \sum_{i=1}^{n} \sum_{j=1}^{c} u_{ij}^{m} \left\| x_i - c_j \right\|^2 \tag{2.3}$$

where,

$$u_{ij}^{m} = \frac{1}{\sum_{k=1}^{c}\left(\frac{\|x_i - c_j\|}{\|x_i - c_k\|}\right)^{2/(m-1)}}, \text{and } c_j = \frac{\sum_{i=1}^{n} u_{ij}^{m}.x_i}{\sum_{i=1}^{n} u_{ij}^{m}} \tag{2.4}$$

From Eqs. (2.3) - (2.4) it can be inferred that the FCM objective function differs from K-means by incorporating membership values u_{ij} for various observations x_i in the feature space as well as a "fuzzifier" term $m \mid m \in \{1 \le m \le \infty\}$ that directs the extent of cluster fuzziness.

In their work, Yang et al. [Yang et al., 2002] proposed two Eigen-based fuzzy clustering routines namely, Separate Eigenspace FCM (SEFCM) and Couple Eigen-based FCM (CEFCM), for segmentation of objects with desired attributes in color imagery. Given an arbitrary image with a predefined set of pixels, the color space in which the image is expressed is initially divided into two eigenspaces called principal and residual eigenspaces using the Principal Component Transformation. Following this, the SEFCM design obtains a segmentation output by integrating the results of independently applying the FCM algorithm to the aforementioned eigenspaces. The integration process involves a logical selection of common pixels from the two clustering results. On the other hand, the CEFCM arrangement obtains an output segmentation result by jointly considering the principal and residual eigenspaces. Both routines were found to outperform the traditional FCM clustering approach from a color object segmentation perspective. Liew et al. [Liew et al., 2005] instituted an adaptive fuzzy clustering scheme by imposing local spatial continuity using contextual information. The method was targeted for exploiting inter-pixel correlation existent in most conventional imagery, in a fuzzy framework. Chen et al. [Chen et al., 2005] proposed a computationally efficient version of the FCM algorithm using a two-phase scheme involving data reduction followed by clustering. This computationally more efficient approach was found to produce results of similar quality to the conventional FCM. More recently, Hung et al. [Hung et al., 2008] developed a Weighted FCM (WFCM) clustering technique wherein the weights for various features were computed using a bootstrap method. Incorporating the bootstrap approach was found to provide satisfactory weights to individual features from a statistical variation viewpoint, and enhance the performance of the WFCM procedure. Tziakos et al. [Tziakos et al., 2009] proposed an approach using the Laplacian Eigen (LE) map algorithm, a manifold learning technique, to boost the performance of FCM clustering. The methodology is commenced by extracting local image characteristics from overlapping regions in a high dimensional feature space, from which a low dimensional manifold was learned using spectral graph theory. Following this, the LE-based dimensionality reduction technique was used to compute a low dimensional map that captured local image characteristic variations, eventually

used to enhance the performance of FCM clustering. Krinidis et al. [Krinidis et al., 2010] and Wang et al. [Wang et al., 2010] developed alternate yet efficient versions of the FCM scheme that employed both local intensity and spatial information. Yu et al. [Yu et al., 2010] founded an Ant Colony–Fuzzy C-means Hybrid Algorithm (AFHA) for color image segmentation that adaptively clustered image elements utilizing intelligent cluster center initializations as well as sub-sampling for computational efficiency. The results of the AFHA structure were found to have smaller distortions and more stable cluster centroids over the conventional FCM.

Besides the practices discussed so far in this section, several unique clustering-based methods for image segmentation have also been proposed. Veenman et al. [Veenman et al., 2003] developed an efficient and optimized model for clustering using a Cellular Co-evolutionary Algorithm (CCA) that does not require any prior knowledge of the number of clusters. On the other hand, Allili et al. [Allili et al., 2010] instituted a clustering model that combined a generalized Gaussian mixture model with pertinent feature selection to alleviate problems of under/over segmentation. Jeon et al. [Jeon et al., 2006] introduced a sparse clustering method for color image data using the principle of Positive Tensor Factorization (PTF). Aghbari et al. [Aghbari et al., 2006] proposed a Hill-Manipulation process where the protocol of segmenting an arbitrary color image was treated in an equivalent fashion to that of finding hills in its corresponding 3D intensity histogram. Ma et al. [Ma et al., 2007] introduced the notion of clustering using lossy data coding and compression, and demonstrated their work on natural scene color images. The algorithm in [Ma et al., 2007] was employed by Yang et al. [Yang et al., 2008], who proposed a Compression-based Texture Merging (CTM) routine that treated segmentation as a task of clustering textural features modeled as a mixture of Gaussian distributions, wherein the clustering methodology was acquired from a lossy data compression protocol. Sample segmentation outcomes of the CTM algorithm using default parameters (coding data length parameter γ =0.2) are exhibited in Fig. 2.6. Huang et al. [Huang et al., 2011] advocated the concept of pure ‘clustering-then-labeling’ for efficient segmentation of color images.

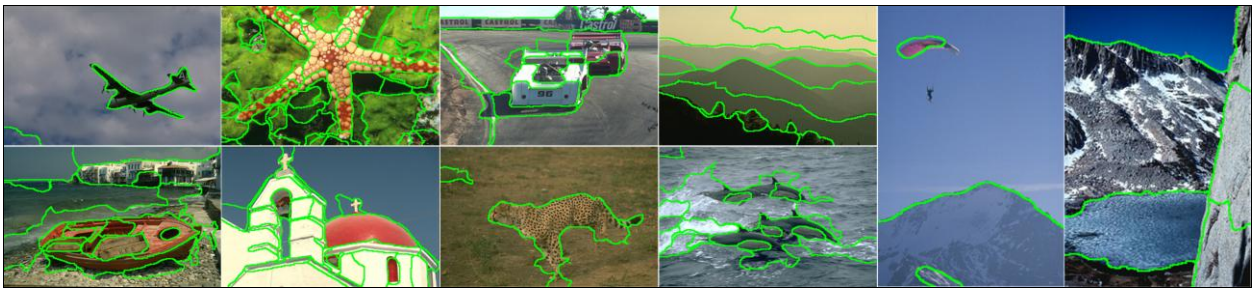

Fig. 2.6: Results of the CTM algorithm [Yang et al., 2008].

2.2.1.1.2 Histogram Thresholding

Histogram thresholding (see [Sezgin et al., 2004] for a comprehensive survey) is a spatially blind technique primarily based on the principle that segments of an image can be identified by delineating peaks, valleys and/or shapes in its corresponding intensity histogram. Similar to clustering, histogram thresholding protocols require minimal effort to realize in comparison to most other segmentation algorithms, and function without the need for any a-priori information about the image being partitioned. Owed to its simplicity, intensity histogram thresholding initially gained popularity for segmenting gray-scale images. However, during its course of development, it was found that thresholding intensity histograms did not work well for low contrast images without obvious peaks, and yielded ambiguous partitions in the presence of spurious peaks manifested by noise. Additionally, for color images, it was determined that thresholding in a multi-dimensional space is a difficult task. Fig. 2.7 illustrates color histograms of the 'starfish' and 'boat' images in the RGB space, generated using an open source ImageJ plugin called Color Inspector 3D [Barthel, 2006]. Each color bin in the 3D histogram is represented as a sphere whose volume is proportional to the frequency of occurrence of that color. From the histograms, it can be observed that finding multiple thresholds to efficiently partition them presents a challenging problem.

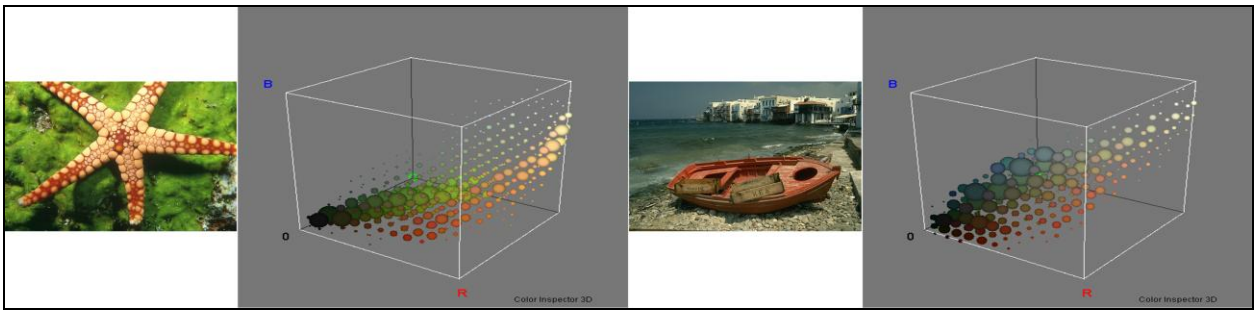

Fig. 2.7: Sample color images with their corresponding 3-D color histograms.

Kurugollu et al. [Kurugollu et al., 2001] proposed an algorithm for color image segmentation that involved two major steps namely multithresholding and fusion. The method is initiated by forming 2D histograms using pair-wise band combinations (RG, GB, and BR), each of which were subjected to a peak finding protocol. Following this, based on the delineated peaks, a multithresholding scheme was used to form a segmentation result unique to each pair of channels. These three segmentation results were fused using a label concordance algorithm and refined using a spatial chromatic majority filter to derive the final segmentation result. In a similar framework, Cheng et al., [Cheng et al., 2002] designed a color image segmentation scheme, based on the idea of thresholding a homogram, that simultaneously captured the occurrence of gray levels along with adjoining homogeneity values among pixels. The segmentation routine was initiated by forming a homogram individually for each color channel, the peaks of which

were used to guide a subsequent thresholding scheme to acquire an initial over-segmented set of clusters. Finally, the three sets of clustering results from the red, green and blue planes respectively were united together to achieve the final segmentation. Mushrif et al. [Mushrif et al., 2008] exploited the concept of Histon thresholding based on rough set theory to devise an efficient algorithm for color image segmentation. A Histon is defined as a set of pixels that are all potentially part of a particular segment. Their three-step architecture involved computing a Histon, followed by thresholding and culminating in a region merging process (discussed in Section 2.2.2.1). Additionally, they further enhanced the aforementioned methodology though the work in [Mushrif et al., 2009] using an Atanassov's Intuitionistic Fuzzy Set (A-IFS) Histon representation of the input image. In their work, Manay et al. [Manay et al., 2003] proposed an adaptive thresholding structure for fast segmentation using an anisotropic diffusion model based on the principle that, for an arbitrary local area, an optimal threshold can be derived close to image edges using a smooth version of it.

2.2.1.2 Spatially Guided Approaches

In contrast to spatially blind methods, spatially guided approaches as the name suggests, are ones that are guided by spatial relationships of pixels for segmentation. Their primary objective is to form pixel groupings that are compact or homogeneous from a spatial standpoint, irrespective of their relationships in specific feature spaces. However, despite the development of many spatially guided techniques the use of region and edge information explicitly or in an integrated framework remain widely- accepted alternatives for the formation of spatially compact regions.

Segmentation techniques that distinctly use region information typically employ protocols involving growing, splitting, and merging individually or in suitable combinations. For the formation of an arbitrary region, growing is a process which starts from a single pixel or small pre-defined labeled set of pixels called a seed, and based on a certain homogeneity criterion iteratively accumulates pixels around it, as depicted in Fig. 2.8. The growth of a region stops when pixels satisfying the homogeneity criterion are no longer found. Most growing approaches help create spatially connected and compact regions relative to other routines in literature. Additionally, the established regions possess specific user defined properties with high tolerance to noise. However, sequential design (pixel by pixel agglomeration) of growing procedures often results in intensive computational schemes with significant memory requirements. In addition, the presence of varying shades of colors produce, in general, over-segmented outputs. Furthermore, the quality of the segmentation is heavily dependent on the order in which the seeds are processed.

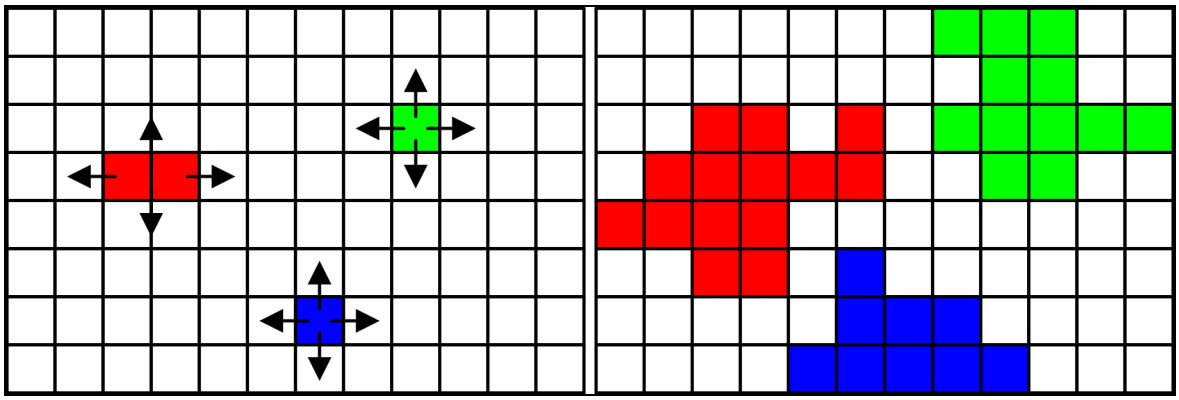

Fig. 2.8: Seed pixels (left), and region formed after a few iterations of growing (right).

In comparison to region growing, region splitting is a technique that is initiated with an inhomogeneous segmentation of an image, which is repetitively split until segments satisfying a particular homogeneity criterion are obtained. Splitting can be achieved via diverse methods such as quadrature tree decomposition, watersheds, or implicitly via region growing when multiple seeds are placed in homogeneous areas that fall under different categories of our low level taxonomy. Consequently, we do not explicitly group them in our discussion. The aforementioned growing and splitting procedures generally yield good results for simple images with well-defined homogeneous regions. However, utilizing them purely based on color homogeneity may, in general, pose difficulties due to varying shades of color, non-uniformity of color spaces, illumination and texture disparities. Region merging is a process in which sub-regions - potentially part of a larger identifiable region - are fused together to yield a reduced set of segments that are spatially meaningful with respect to the input image content (see Fig. 2.9 for a simplified illustration). In general, for reasonably complex images, growing/splitting methods often result in an over-segmented region map. As a result, they are often integrated with some type of a region merging scheme to improve the final outcome.

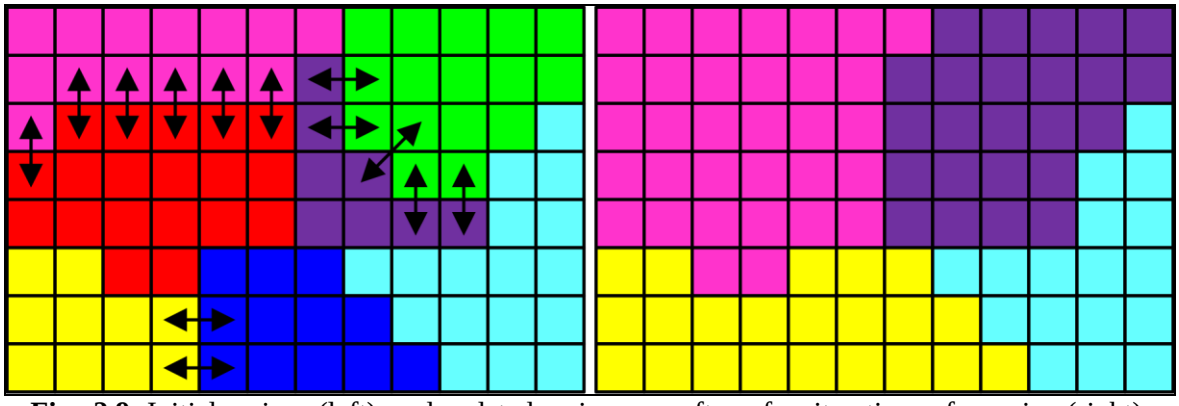

Fig. 2.9: Initial regions (left), and updated region map after a few iterations of merging (right).

2.2.1.2.1 Region-Growing Approaches

Fan et al. [Fan et al., 2001] proposed an automatic image segmentation algorithm that begins with an edge detection scheme, wherein the centroids between the detected edges are chosen as the set of candidate seed points. Subsequently, a growth procedure is utilized to spatially integrate pixels, in a recursive fashion, to an appropriately chosen seed from the entire set until the final segmentation is achieved. Wan et al. [Wan et al., 2003] were the first to introduce a theoretical criteria for a specific category of region growing algorithms called symmetric region growing, that are insensitive to the selection of the initial seed points. Fondón et al. [Fondón et al., 2006] introduced a multistep region growing procedure for color image segmentation, in which the extent of growth can be controlled using a tolerance parameter dependent on the variance of the actual grown region. Although the method was successful in accurately delineating spatial extent of regions, it necessitated manual selection of seed points and could only handle one region at a time. Qin et al. [Qin et al., 2010] advocated the use of semantic information for effective region growing, and proposed an MRF-based multivariate image segmentation algorithm.

2.2.1.2.2 Region-Merging Approaches

Similar to growing, a significant number of approaches have been proposed that explicitly use a merging protocol for region-based segmentation. Devaux et al. [Devaux et al., 2001] built a unique segmentation architecture that employed the Karhunen-Loeve transform (KLT) in combination with color and textural attributes for region-based segmentation of color aerial images. The algorithm separately exploited color and texture information to come up with two initial segmentation maps that are subsequently fused together in a merging protocol. Chen et al. [Chen et al., 2004] developed a segmentation technique based on color contrast. The technique began by converting the color input image from RGB to CIE L*a*b* and then utilized the later computed values to estimate contrast information with four directional operators. The estimated contrast map was thresholded to identify an initial set of regions, which were appropriately merged using a connection and verification process. Nock et al. [Nock et al., 2004] explored a statistical region merging structure of segmenting image data, based on the notion that perceptual grouping with region merging can be effectively used to encapsulate the big picture of a scene, using primary local glimpses of it. The work in [Nock et al., 2004] was further enhanced in [Nock et al., 2005] by treating statistical region merging as a non-parametric mixture model estimation problem. In his work, Kim [Kim, 2005] devised an approach for segmenting low-depth-of-field images using morphological filters and region merging. The procedure involved an initial conversion of a low-depth-of-field image to an alternate feature space representing Higher Order Statistics (HOS). The resultant HOS map was simplified using morphological reconstruction followed by region merging to produce the output segmentation result. Kuan et al. [Kuan et al., 2008] presented a novel region merging strategy for

segmenting salient regions in color images. The technique generated an initial set of regions by extracting dominant colors in the input image, using a non-parametric density estimation methodology. Subsequently, a merging protocol based on "importance index" and "merging likelihood" criterions was used to refine the initial set. With a similar global objective to the work in [Kuan et al., 2008], Liu et al. [Liu et al., 2011] proposed an unsupervised segmentation algorithm aimed at salient object extraction. The method was based on region merging in a Binary Partition Tree (BPT) framework. It utilized a novel dissimilarity measure that considered color difference, area factor, and adjacency degree criterions. A robust termination criterion for conventional region merging algorithms based on a novel distinctness predicate of adjacent regions was proposed in [Tan et al., 2009]. The effectiveness of the aforementioned criterion was demonstrated using two new merging criteria based on local and global image characteristics. Region merging techniques using statistical measures from the field of information theory was first seen in the work of Calderero et al. [Calderero et al., 2010]. The proposed merging protocols were unique in the fact that they did not make any assumptions of color or texture homogeneity of regions, but were characterized by non-parametric region models.

2.2.1.2.3 Hybrid Growing-Merging Approaches

Integration of growing and merging is another popular region-based methodology in the segmentation realm. Deng et al. [Deng et al., 2001] proposed the prominently known J-SEGmentation (JSEG) algorithm that integrated color quantization and spatial segmentation for extraction of color-texture regions in images and video (see Fig. 2.10). The JSEG method commences in a color quantization step utilized to obtain a 'color class map' which is subsequently employed to compute a 'J-image' based on certain spatial constraints. These spatial constraints were designed such that the resultant J-image corresponded to measurements of local homogeneities that acquired high values at region boundaries and low values for homogeneous color-texture regions. Subsequently, the J-image is utilized as a reference to identify suitable seed points to initiate a region growing process, wherein the obtained regions are eventually refined in a merging process using a color similarity criterion. Although the JSEG method was efficient in deriving spatially compact regions, it suffered from the fact that the use of color quantization caused over-segmentation in regions of varying shades due to illumination or texture disparities, as viewed in some of the results of Fig. 2.10 (see cheetah, skydiver and lake images). The segmentation outcomes displayed in Fig. 2.10, were obtained using default parametric settings, where the parameters named color quantization threshold ($qthresh$) and number of scales ($Iscale$) are, by default, automatically computed, while the region merge threshold parameter ($mthresh$) was set equivalent to 0.4.

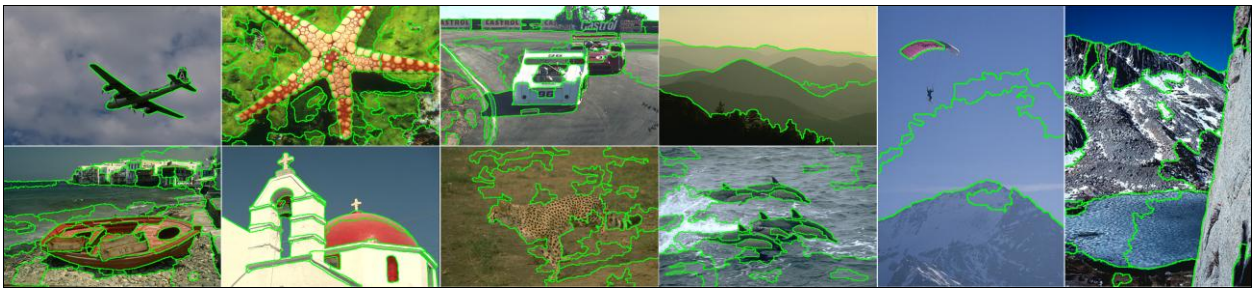

Fig. 2.10: Results of the JSEG algorithm [Deng et al., 2001].

The aforestated drawback of the JSEG technique to a certain extent was addressed by Wang et al. [Wang et al., 2006], who advocated the use of mean shift clustering instead of color quantization for improved results. In their work, Wang et al. [Wang et al., 2006] uncovered another drawback of the JSEG procedure by demonstrating that ignoring color discontinuity in the computation of the J-measure caused over-segmented results. To overcome this deficiency, they proposed a novel hybrid measure for homogeneity. Amidst other hybrid approaches, Cheng [Cheng, 2003] postulated a segmentation procedure for color image data in a growing-merging framework integrated with 3-D clustering and relaxation labeling. Shih et al. [Shih et al., 2005] developed a segmentation algorithm based on seeded region growing and merging, incorporating strategies to avoid pixel order dependencies. He et al. [He et al., 2007] employed the concept of Gradient Vector Flow (GVF) in a seeded region growing and Region Adjacency Graph (RAG)-based merging architecture. Dynamic color gradient thresholding (DCGT) integrated with a growing-merging scheme was first seen in the work by Balasubramanian et al. [Balasubramanian et al., 2008]. The DCGT technique was used to guide a region growth procedure for the formation of an initial set of regions which were further refined in a merging protocol. Both steps were performed purely based on color information. To this effect, the DCGT algorithm faced problems of over segmentation due to the lack of a texture descriptor and was computationally expensive. Fig. 2.11 portrays the segmentation outcomes achieved from the DCGT algorithm using default parametric values described in [Balasubramanian et al., 2008].

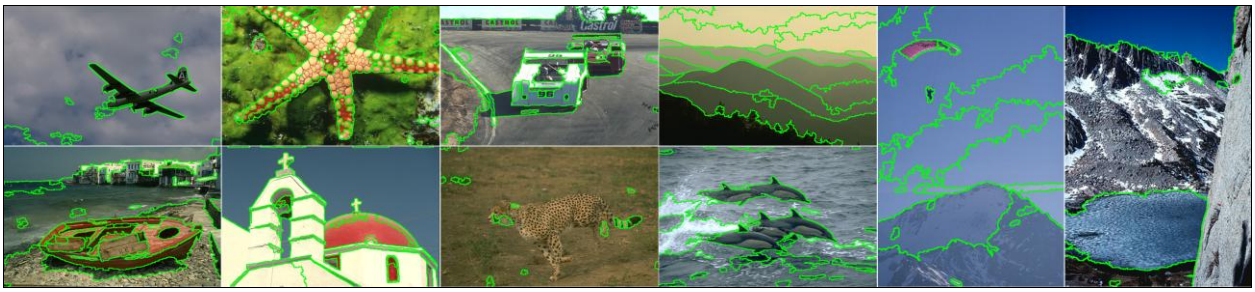

Fig. 2.11: Results of the DCGT algorithm [Balasubramanian et al., 2008].

More recently, Krinidis et al. [Krinidis et al., 2009] instituted an approach for color texture image segmentation in a growing-merging schema based on a 3D physics-based deformable surface model

derived from intensity and spatial information of images. Color image segmentation using the Dual Tree Complex Wavelet Transform (DT-CWT) integrated with a growing-merging strategy was seen in [Celik et al., 2010]. The partitioning process was initiated by the DT-CWT computation that enabled multiscale edge detection, wherein the acquired edges were subjected to binary morphological operations to locate suitable seed points. These seed points were employed in a region growing approach to delineate an initial set of regions, which were fine-tuned in a subsequent merging process. Recently, Panagiotakis et al. [Panagiotakis et al., 2011] devised a scheme for natural image segmentation in a growing-merging structure based on tree equipartion and Bayesian flooding processes for feature extraction. Additionally, several hybrid region-based approaches [Sclaroff et al., 2001, Gevers, 2002, Navon et al., 2005, Luo et al., 2006, Prasad et al., 2006, Lézoray et al., 2009, Wan et al., 2011, Sobieranski et al., 2011] have also been proposed.

In contrast to the segmentation approaches discussed in the last three sub-sections, energy-based segmentation techniques aim to minimize explicit cost functions. They can, in general, be classified into ones that explicitly utilize edge/contour-based energy (eg. active contours) or ones that employ region-based energy to delineate different regions (eg. Mumford-Shah formulation and Bayesian techniques like Markov Random Fields (MRFs)).

2.2.1.2.4 Active Contours

Within the notion of using edge/contour-based energy, curve evolution methods involving active contours better known as 'evolving fronts' have gained tremendous popularity over the last decade. From a high level view-point, active contours can be categorized based on their implementation as being either Parametric Active Contours (PACs) or Geometric Active Contours (GACs).

PACs are generally represented in a Lagrangian formulation where the evolving curves are called 'snakes', a concept first introduced by Kass et al. [Kass et al., 1988]. A snake is defined as a curve or a deformable spline $v(s) = (x(s), y(s))$ that constantly moves/evolves based on a specific energy model $E(v)$ until it attains a shape that best fits an object (or multiple objects) of interest in the scene. This energy functional typically comprises of internal ($E_{\text{int}}(v(s))$) and external ($E_{ext}(v(s))$) energy terms as shown in Eqs. (2.5) - (2.6), whose combined effect drives a snake towards the boundary of an object resulting in the overall energy being minimized, given as:

$$E(v) = \int_0^1 \left[E_{\text{int}}(v(s)) + E_{ext}(v(s)) \right] ds \tag{2.5}$$

$$E(v) = \frac{1}{2} \int_0^1 \left[\alpha(s) \left| \frac{\partial v(z)}{\partial s} \right|^2 + \beta(s) \left| \frac{\partial^2 v(z)}{\partial s^2} \right|^2 \right] ds + \int_0^1 E_{ext}(v(s)) ds \tag{2.6}$$

In the aforementioned equations, (x, y) symbolizes the co-ordinates of a snake in the image domain, while s is proportional to its arc length. Furthermore, $E_{\text{int}}(v(s))$ is contour dependent. It is utilized to control its tension and rigidity via parameters $\alpha(s)$, $\beta(s)$ respectively, and is minimized when a snake possesses a shape that is in close proximity to the object of interest. On the other hand, $E_{ext}(v(s))$ is explicitly calculated in the image domain, and is minimized when the physical location of a snake is along the boundaries of the object of interest. Among PACs, there exists a class of snakes called region-based active contours given that they are designed to attract to boundaries that distinguish homogeneous regions. Since its inception, it has been uncovered that the traditional snake model suffers from two major drawbacks that derail it from converging on the desired object of interest. The first occurs when the contour initialization is far from the true object boundary and, the second is when the object of interest has cavities that are concave in nature. To overcome the first shortcoming, multiresolution methods, pressure forces, as well as several enhanced models such as balloon/distance snake models have been proposed. On the other hand, methods involving Gradient Vector Flow (GVF) and directional snake models have been offered to account for the second deficiency. PACs have several merits over classical segmentation techniques such as: 1) they are self-accommodative in their pursuit for a global energy minimum, 2) they can be easily molded via the $E_{ext}(v(s))$ term as needed, 3) they can be designed to be image scale dependent, and finally 4) they are not biased towards any particular object/region shape and consequently are very effective for segmenting/tracking objects in spatio-temporal dimensions. Major potential demerits of PACs include: 1) brazing localized energy minima, 2) ignoring minor image features for global energy minimization, 3) focusing only a few regions at a time, and 4) relying on stringent convergence criteria for high accuracy. Dumitras et al. [Dumitras et al., 2001] proposed a three-step algorithm using angular-map driven snakes for shape description and extraction of objects in color imagery. The first step involved computation of an angular map using all color pixel vectors and a reference vector that characterizes color changes in the input image. This map is utilized as input to an edge detection protocol in the second stage of processing. Finally, the resultant edge map is presented to a snake model to segment the object of interest. Dumitras et al. experimented with distance and GVF snake models in their work. The use of PAC evolution based on a cubic smoothing spline for real time segmentation of images was first seen in the work of Precioso et al. [Precioso et al., 2005]. Moreover, through this work Precioso et al. demonstrated that the choice of a smoothing spline approximation instead of spline interpolation makes a snake more robust to noise variations. More recently, Ozertem et al. [Ozertem et al., 2007] introduced a non-parametric formulation of a snake energy function using kernel density estimation that exploited the underlying kernel density estimate of the image data. Lankton

et al. [Lankton et al., 2008] propounded a method on region-based active contours driven by localized region energy calculations for improved segmentation accuracy.

In comparison to PACs, GACs are implicitly represented in an Eulerian formulation where evolving curves are evaluated as the level sets of a distance function in two-dimensions; a theory first introduced for image segmentation by Malladi et al. [Malladi et al., 1995] based on the work originally done by Osher et al. [Osher et al., 1988]. The key idea of a level set-based segmentation method is to commence with a closed contour Γ in two dimensions, which is eventually made to propagate in a direction orthogonal to itself at a specific speed F, driven by a higher dimensional scalar function defined over the input image. Thus, the evolving front at any location (x, y) is derived as the zero level set of the aforestated scalar function at time instant t, mathematically represented as:

$$\Gamma = \{\Phi(x, y, t) : \Phi(x, y, t) = 0\} \tag{2.7}$$

Employing the chain rule on Eq. (2.7) and performing specific algebraic simplifications, the evolution of Φ (given the value of $\Phi(x, y, t) = 0$) can be expressed as:

$$\Phi_t + \|\nabla\Phi\| F = 0 \tag{2.8}$$

Eq. (2.8) is popularly referred to as the level set equation, and serves as a useful tool to track the evolution of contours along images. Fig. 2.12 shows the segmentation results obtained using an open source Level set tool box [Sumengen, 2005] using default parametric settings. The primary virtue of GACs over alternate contour energy based approaches is that its implicit boundary formulation can efficiently undergo topological changes pertinent to splitting or merging. Consequently GACs are better suited for shape-invariant multi-region/object segmentation. A secondary asset of GACs over conventional schemes is its non-parametric nature that allows it to be generically used for disparate datasets.

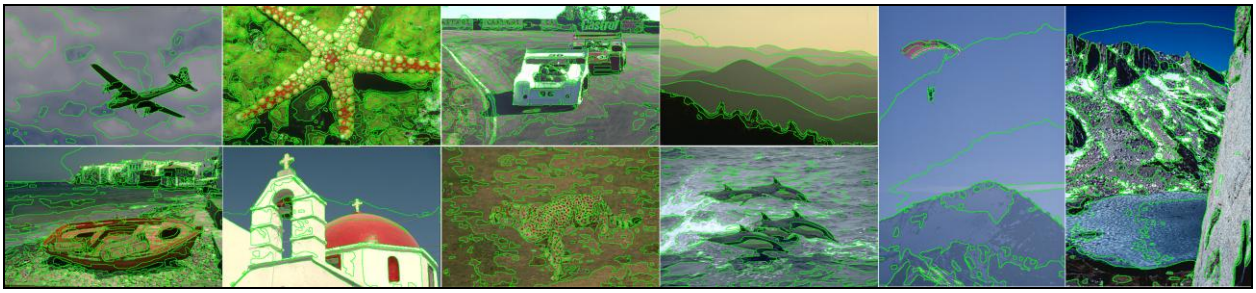

Fig. 2.12: Level set segmentation results obtained using an open source toolbox [Sumengen, 2005].

Brox et al. [Brox et al., 2006] designed a GAC-based segmentation approach for multiple regions utilizing coupled level set functions. More specifically, their segmentation framework employed one level set function for each region. It was novel because it allowed for the extraction of an arbitrary number of regions unlike conventional level set approaches that optimally extracted one or two regions. Michailovich et al. [Michailovich et al., 2007] proposed a segmentation method for gray and color images

based on GACs using an energy function that incorporated the Bhattacharya distance. The underlying algorithm was based on the notion that regions in an image can be delineated using a curve evolved in a manner such that the Bhattacharya distance between the estimates of probability densities among various segmentation classes is maximized. To this effect, a cost function that measured the dissimilarity between contents of regions was defined, and employed such that a contour was made to converge to a shape that minimized overlap between the aforementioned contents (or equivalently maximized the distance among probability densities of various segments). Another approach targeted at segmenting multiple regions in color images using GACs was demonstrated in the work of Ayed et al. [Ayed et al., 2008]. The proposed technique allowed the number of regions being segmented to be automatically varied via a region merging methodology Furthermore, Bertelli et al. [Bertelli et al., 2008] employed GACs that evolved based on cost functions derived from within-region (pixel-pair) dissimilarity instead of between-region cuts in a binary or multiphase level set composition, for multi-region segmentation of color images. Xie [Xie, 2008] introduced a GAC model that incorporated magnetostatic interactions among object boundaries and active contours for efficient segmentation. The proposed approach was found to be extremely robust in the presence of complex geometries as well as problems pertinent to contour initialization and weak/broken edges. The algorithm in [Xie, 2008] was enhanced in [Xie, 2010] with and without random initializations of active contours, for object segmentation in color imagery. Fuzzy energy-based active contours using a pseudo-level set formulation for object detection/segmentation in color images was seen in the work by Krinidis et al. [Krinidis et al., 2009]. The advantages of their approach were, in general, two-fold: 1) The underlying cost function was capable of detecting objects whose boundaries were not well defined by gradient calculations, and 2) The fuzzy energy alteration to the conventional model enabled a contour to quickly converge to the desired object of interest within a few iterations. Li et al. [Li et al., 2010] instituted a unique segmentation algorithm using GACs in a level set formulation wherein the regularity of the level set function was maintained during the curve evolution process. The approach was designed to minimize a cost function derived from gradient flow, and comprised of a distance regularization term in conjunction with an external energy term that forced the zero level set to the desired locations. On the other hand, Salah et al. [Salah et al., 2010] introduced a process for multi-region segmentation of image data in a level set framework, using an energy functional that encompassed a term to evaluate the deviation of mapped data (acquired using a kernel function) within each segmented region from a piecewise model, in addition to a classic regularization term that enforced smoothness of region borders. More recently, Karoui et al. [Karoui et al., 2010] proposed an algorithm for segmentation of textured regions in images using a GAC-based level set formulation, which minimized a cost function comprised of a similarity measure between region features and texture descriptors, along with a boundary-based component that imposed smoothness/regularity of region

boundaries on the evolution process. Ghosh et al. [Ghosh et al., 2010] pioneered a single object segmentation algorithm in a variational formulation based on edge flow vectors derived from several image features pertaining to color, texture and intensity edges. Wang et al. [Wang et al., 2011] devised a color segmentation protocol based on the work in [Deng et al., 2001] using a level set formulation that minimized a global inhomogeneity metric for segmentation of photographic imagery.

In context of GAC's-based curve evolution, other notable contributions involved the use of Geodesic active contours for region/object segmentation. Geodesic active contours are dynamically modeled level set methods that facilitate combining common curve evolution practices with energy minimization techniques, and are considered as the geometric alternative for snakes. In their work, Goldenberg et al. [Goldenberg et al., 2001] devised a computationally efficient implementation of a geodesic active contour model that was numerically consistent using a narrow band level set formulation and a fast marching technique. Their implementation, involving advanced numerical methods, was found to efficiently solve a geometric non-linear model for applications involving region segmentation and object tracking. Furthermore, Sagiv et al. [Sagiv et al., 2006] utilized a geodesic active contour framework for segmenting textured regions in natural scene images, where the texture was modeled using Gabor filters sensitive to a set of orientations, scales and frequencies.

2.2.1.2.5 Mumford-Shah Functional-based Approaches

Chan et al. [Chan et al., 2001] proposed a unique model for active contour-based segmentation using the Mumford-Shah (MS) functional. If $\vec{C}$, g and f represent a smooth closed segmenting contour, the observed image data and its piece-wise approximation respectively, then the Mumford-Shah energy functional $E(f,\vec{C})$ is defined as:

$$E(f,\vec{C}) = \beta\iint_{\Omega}(f-g)^2 dA + \alpha\iint_{\Omega\setminus\vec{C}}\left|\nabla f\right|^2 dA + \gamma\oint_{\vec{C}} ds \tag{2.9}$$

In Eq. (2.9) Ω symbolizes the image domain while the α, β, γ parameters appropriately weight various terms in the functional in order to control the segmentation scale and the smoothness of the outcome. The use of the MS functional enabled the development of a curve evolution process for region/object segmentation without utilizing edge information as a stopping criterion, as typically employed by conventional approaches. Tsai et al. [Tsai et al., 2001] extended the work of Chan et al. [Chan et al., 2001] by divulging an estimation-theoretic approach to curve evolution using the MS functional for color image segmentation and smoothing. Heiler et al. [Heiler et al., 2005] proposed an algorithm for segmentation based on integrating the use of natural image statistics with the technique in [Chan et al., 2001]. In their work, Gao et al. [Gao et al., 2005] demonstrated a new hierarchical image

segmentation and smoothing algorithm based on a multiphase level set formulation and the Chan-Vese piecewise constant smooth active contour model.

2.2.1.2.6 Bayesian Segmentation Techniques

A class of energy-based segmentation approaches that have maintained continued interest amongst researchers over the past few years stems from the field of probability theory wherein region characteristics are modeled through Markov Random Fields (MRFs). In a standard MRF-based procedure the segmentation objective is formulated using the Maximum A-posteriori Probability (MAP) criterion. According to the MAP criterion, a desired segmentation $\hat{S}$ is defined as the one that maximizes the a posteriori probability $p(S/I)$ of segmentation S, pertaining to an observed image I, which according to the Bayes rule is mathematically expressed as:

$$p(S/I) = \underset{S\in\Omega}{argmax}\{p(I/S)p(S)\} \tag{2.10}$$

The terms $p(I/S)$ and $p(S)$ are known as the class conditional and a-priori probability distributions respectively, while Ω represents the set of all possible segmentation outcomes. The conditional probability is typically responsible for characterizing the underlying distribution of intensity values or other attributes/features in an image. In contrast, the a-priori probability distribution is employed for imposing spatial connectivity constraints for region formation. Several optimization approaches (e.g. Iterative conditional modes (ICM), Highest Confidence First (HCF), and simulated annealing) integrated with parametric estimation methods (e.g. Expectation Maximization (EM), Maximum Likelihood (ML) estimation) have been utilized to achieve the aforementioned maximization. The primary advantage of Bayesian-based segmentation is its mathematically principled approach that uses statistical inference and a-priori information about the underlying data. Consequently it is extensively favored for segmenting images that contain non-deterministic content such as textures and statistical noise, which often prove challenging for traditional segmentation modus operandi.

In his work, Mukherjee [Mukherjee, 2002] demonstrated an MRF-based algorithm that used a multidimensional luminance-chrominance feature space for improved segmentation of color images. The initial segmentation estimate for the MRF model was obtained using a conventional region growing approach and the resultant regions were refined using a merging protocol. Gao et al. [Gao et al., 2002] described a color image segmentation technique involving color conversion and MRF-EM modeling of regions in a unique "narrow band" multiresolution implementation. The algorithm was initialized by a transformation of the input image from RGB to LUV for improved color differentiation which was subsequently utilized as input to an MRF processing module. The parameters of the MRF model were

estimated through an EM technique, wherein MRF-EM processing was invoked in a multiresolution framework for computational efficiency. Luo et al. [Luo et al., 2003] introduced the concept of non-purposive grouping (NPG) that defined the expectations of a perceptually acceptable segmentation outcome, and proposed a probabilistic model to the NPG problem using an MRF formulation of regions in a HCF framework. Deng et al. [Deng et al., 2004] evinced a function-based weighting parameter between the class conditional and the a-priori probability components of an MRF model, for image segmentation. The aforesaid weighting parameter was utilized to overcome the training phase typically necessitated to estimate MRF model parameters, consequently making the algorithm completely unsupervised. Tab et al. [Tab et al., 2006] came up with a multiresolution color image segmentation approach that was capable of segmenting regions with similar patterns across different resolution levels, by incorporating MRF modeling of regions and the discrete wavelet transform. The algorithm was specifically targeted at object-based wavelet coding applications. Xia et al. [Xia et al., 2006] instituted an adaptive algorithm for segmenting textured images using a couple Markov Random Field (CMRF) comprised of a finite symmetric conditional Markov model (FSCM) that typified an image for feature extraction/estimation, and a multilevel Logistic model (MLL) that characterized the labeling process to arrive at the desired segmentation outcome. The FSCM and MLL models were mutually dependent terms that were implemented using a simulated annealing scheme. Kato et al. [Kato et al., 2006] pioneered an ICM-based architecture MRF model, which combined color and texture information for segmentation. Moreover, to facilitate accurate estimation of parameters required by the model, an estimation methodology based on the EM algorithm was developed. A couple of years later, Kato [Kato, 2008] proposed a color image segmentation algorithm based on an MRF formulation in which pixel classes were characterized by a multivariate Gaussian distribution. Estimates of the number of classes, class model parameters and pixel labels that made up the renowned "incomplete data problem" were all derived from a posterior distribution using a reversible jump Markov Chain Monte Carlo (RJMCMC) method. Diplaros et al. [Diplaros et al., 2007] presented a novel generative model and EM algorithm for MRF-based color image segmentation that generated a class label for a pixel using prior distributions that shared similar parameters with its neighboring pixels. The proposed algorithm was found to possess an advantage of being relatively simple from an implementation standpoint with competing qualitative performance against traditional MRF-based segmentation methods. Nikou et al. [Nikou et al., 2007] proposed a new family of Gauss-Markov random field (GMRF)-based smoothness priors for modeling class/label probability distributions needed in a conventional MRF using spatially variant finite mixture models (SVFMMs) for color image segmentation. An important aspect of the use of GMRFs in this work was that it took into account individual class statistics to enforce class-dependent smoothness constraints. More recently, the work in [Nikou et al., 2007] was further used by Nikou et al. [Nikou et al., 2010] to

illustrate an advanced hierarchical Bayesian model for mixture model-based segmentation. Mignotte [Mignotte, 2010] designed an MRF fusion model targeted at quickly integrating results estimated from multiple relatively simplistic clustering models to eventually yield an enhanced/accurate final segmentation. This work was novel in the fact that the MRF fusion model made use of a segmentation evaluation metric called the Probabilistic Rand Index (PRI), wherein to have perceptual significance the fusion was achieved in the penalized maximum PRI sense. Chen et al. [Chen et al., 2010] established a new segmentation algorithm formulated as a labeling problem using a probability maximization model, founded on an iterative optimization scheme that alternated between MAP and ML estimations. This MAP-ML-based technique was shown to qualitatively and quantitatively outperform state of the art segmentation approaches. In their work, Vantaram et al. [Vantaram et al., 2011] furnished a hybrid image segmentation algorithm using a Gibbs Random Field (GRF) model (which is an MRF formulation under specific constraints) to form an initial estimate of the label field. This estimate was subsequently refined using color, texture and gradient features integrated with a split-and-merge mechanism to arrive at the final set of segmented regions (see Fig. 2.13 for sample outcomes using a default minimum segment size parameter S =30 pixels).

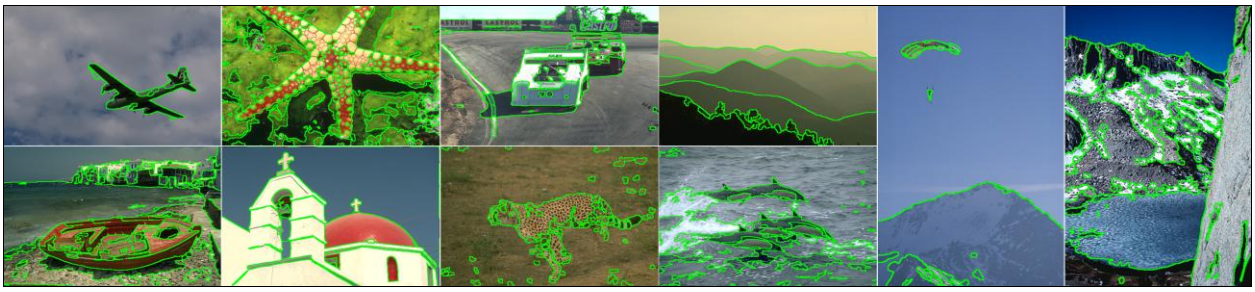

Fig. 2.13: Results of the GRF-based segmentation algorithm in [Vantaram et al., 2011].

Conditional Random Field (CRF), which is an extension of the conventional MRF model, is another probabilistic model that has been effectively used for segmentation. Zhang et al. [Zhang et al., 2010] devised an approach that combined a CRF to fashion spatial relationships among image superpixels with a multilayer Bayesian Network (BN) that sculpted casual dependencies pervasive among different image entities such as regions, edges vertices, to formulate a unified probabilistic graphical model for image segmentation. This unified graphical model was found to surpass the results obtained from prior art that explicitly employed either a CRF or BN for segmentation. On the other hand, Lee et al. [Lee et al., 2010] proposed the use of the AdaBoost machine learning algorithm for identifying disparities between image regions in a CRF formulation, for efficient segmentation in automatic/semi-automatic configurations. The approach was successfully demonstrated on document and natural scene type imagery.

Similar to MRFs, a popular category of Bayesian segmentation methods that originate from the field of probability theory are the ones based on Gaussian mixture models (GMMs) for representing region

processes in images. Carson et al. [Carson et al., 2002] proposed an image segmentation protocol wherein the joint distribution of color, texture and position features were modeled as a mixture of Gaussians whose parameters were estimated using the EM algorithm. The resultant set of regions which Carson et al. called a "Blobworld" representation were employed in an image querying application with enhanced performance over classical content based image retrieval procedures. In a similar architecture, Khan et al. [Khan et al., 2009] presented an unsupervised color image segmentation algorithm where the joint distribution of pixel features pertinent to color, texture, brightness, and position were represented as GMMs. The underlying parameters for the models were assessed using three flavors of the EM algorithm namely penalized EM (PEM), penalized stochastic EM (PSEM) and a novel penalized inverse EM (PIEM) technique. The proposed PIEM method was tested on the Berkeley segmentation database with favorable performance. Integration of the mean shift algorithm and GMMs was first seen in the work by Park et al. [Park et al., 2009], who developed a segmentation technique in which the number of mixture components were estimated using an adaptive mean shift methodology with parameters derived using a mean field annealing EM protocol. The aforementioned mean shift-GMM combination was found to be a competent solution for automatically segmenting color imagery without over-segmentation or isolated region formation. Greggio et al. [Greggio et al., 2011] instituted a fast GMM (FGMM)-based segmentation protocol that automatically inferred the number of components of a GMM as well as their corresponding means and covariances, without necessitating any prior knowledge or conscientious initialization. In contrast to some of the above described GMM-based approaches, Liu et al. [Liu et al., 2011] advocated the use of non-parametric mixture models with multivariate orthogonal polynomials to overcome the dependency of parametric models on a-priori assumptions, for color image segmentation. This methodology, named as the stochastic non-parametric EM (SNEM) technique, was evaluated on the Berkeley database and found to perform well in several empirical situations.

Besides the energy-based techniques discussed in the last three subsections, several other approaches involving specialized energy classes [Arbeláez et al., 2004], energy functionals [Jermyn et al., 2001, Schoenemann et al., 2011], Total Variation (TV)-based functionals [Unger et al., 2008, Donoser et al., 2009], convex relaxation procedures [Pock et al., 2009, Klodt et al., 2011, Brown et al., 2012], curve evolution-based approaches [Allili et al., 2007, Adam et al., 2009, Kokkinos et al., 2009, Li et al., 2009], Bayesian principles [Tu et al., 2002, Ko et al., 2005, Orbanz et al., 2008, Scarpa et al., 2010, Zhang et al., 2011] and other PDE as well as Anisotropic Diffusion-based techniques [Petrovic et al., 2004, Dong et al., 2006, Sofou et al., 2008, Kay et al., 2009] have been developed for various applications.

2.2.1.2.7 Graph-based Segmentation Techniques

Within the realms of energy-driven approaches, there exists a prominent category of techniques that employ graph representations for image segmentation. In a graph theoretic formulation (see Fig. 2.14), an image can be represented as an undirected graph $G = (V, E)$, where every node $v_i \in V$ corresponds to an individual pixel and every edge component $(v_i, v_j) \in E$ pairwise connects neighboring pixel elements in G, represented as vertices v_i and v_j. Additionally, each edge is assigned a weight (w_{ij}) based on the amount of similarity between the two neighboring elements. Thus for an arbitrary image embodied as a graph, the ultimate goal from a segmentation viewpoint is to partition it utilizing metrics that yield a set of disconnected sub-graphs exemplifying meaningful regions that concur with scene content. One significant benefit of graph-based approaches is their capitalization of various cost functions for global energy minimization that more often than not yield optimized segmentation outcomes in a generic framework. On the other hand, graph partitioning methods suffer from significant computational complexity thereby requiring various restrictions and simplifications in order to yield reasonable results in practical real life applications at the expense of the quality of the segmentation.

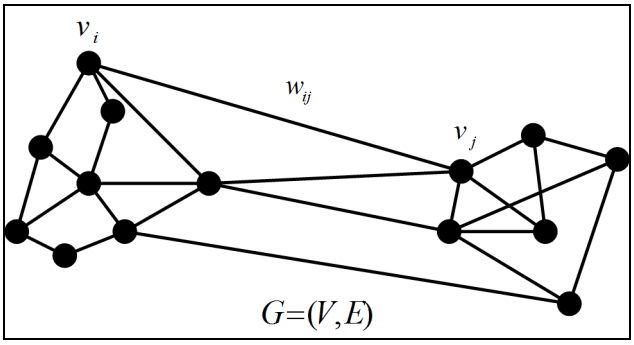

Fig. 2.14: Graph theoretic formulation of an image.

A popular criterion that involves identifying a set of edges crossing a specified path in a graph whose removal results in several disjoint sub-graphs is generically called a 'cut'. Furthermore, the aggregate weight of all the aforementioned edges that cross the cut is called its cost or capacity. It is imperative to understand that segmentation protocols using graphs are predominantly pair-wise graph partitioning processes that begin by splitting G into two partitions and continue to recursively split subsequent subgroups until certain stopping criteria are met. Wu et al. [Wu et al., 1993] was the first to demonstrate a novel spectral graph theoretic approach based on the minimum cut criterion for image segmentation. The algorithm was initiated by the construction of an adjacency graph G analogous to the above described architecture. Moreover, the weights (w_{ij}) of the edge components E were derived using a local derivative operator wherein large discontinuities (or strong edges) were associated with small costs while small discontinuities (or weak edges) were tagged with large costs. Subsequently, the proposed

methodology determined the minimum cuts in the graph G that collectively incurred the least cost during the partitioning process, and in doing so identified closed contours comprised of strong edges in the scene as the final set of segments. A drawback of the minimum cut criterion was that it recurrently yielded small partitions containing isolated nodes in G. To overcome this deficiency, Shi et al. [Shi et al., 2000] proposed the normalized cut (Ncut) measure which is an unbiased metric of disassociation between graph partitions. The Ncut criterion was computed as the cumulative ratio of, the aggregate weight of edge components crossing a cut to the aggregate weight of edge components in various subgroups (that are delineated by a cut) connecting to all nodes in G. To this effect, the Ncut value for isolated nodes was bound to be large, since the aggregate weight of edge components crossing a cut represents a large percentage of the total connections from that small set of isolated nodes to all other nodes in the graph. The introduction of the Ncut standard led to the development of several segmentation approaches [Malik et al., 2001, Tao et al., 2007]. Malik et al. [Malik et al., 2001] devised a segmentation algorithm in a normalized cuts framework that incorporated contour cues with texture features based on the concept of textons, to facilitate formation of regions of consistent brightness and texture. On the other hand, Tao et al. [Tao et al., 2007] established a robust color image segmentation algorithm that integrated the strengths of the mean shift and normalized cuts methodologies with real time performance.

In contrast to the Ncut solution that minimized the total linkage between partitions normalized by the association of the nodes within them, Sarkar et al. [Sarkar et al., 2000] designed a segmentation algorithm based on an average cut metric that minimized the total link weight between partitions normalized by the size of the partitions. Sarkar et al. found the qualitative performance of the average cut criterion from a segmentation perspective to be similar to the Ncut measure with significantly reduced computational load. In their work, Gdalyahu et al. [Gdalyahu et al., 2001] introduced a graph-based segmentation approach by defining a low complexity typical cut criterion robust to noise and spurious cluster formation. The algorithm is initiated by the formulation of samples cuts in G generated using the Karger's contraction algorithm, to eventually derive an average solution called a typical cut for the optimal partitioning of G. The work was demonstrated both on synthetic/real color and gray scale images. Wang et al. [Wang et al., 2001] introduced an image segmentation algorithm that employed a cost function aimed at finding a cut with the minimum mean edge weight in a graph. The mean cut measure possessed several advantages over other graph partitioning approaches as it facilitated cuts with open/closed contours, guaranteed connected partitions and was not influenced by large foreground regions, smooth/short boundaries and similar-weight partitions. A couple of years later, Wang et al. [Wang et al., 2003] generalized the work in [Wang et al., 2001] to define a revolutionary cost function for graph-based segmentation called the ratio cut that enabled region as well as pixel-based segmentation independent of size, shape, smoothness and boundary length attributes. Recently, Kim et al. [Kim et al., 2009] proposed a novel approach for

segmentation of color textured images by formulating the segmentation task as a minimum cut problem in a weighted graph, wherein information from color and texture features were fused in a multivariate mixture model. To find globally optimal minimum cuts, the unsupervised algorithm relied heavily on specific type of split moves.

The graph-based mechanisms discussed thus far are fully automatic and require, in general, substantial computations. On the contrary, a significant effort has been devoted to the development of supervised techniques involving human interaction, to overcome the computational burden manifested by their unsupervised counterparts as well as for achieving results that are more tailored towards user requirements. Boykov et al. [Boykov et al., 2001, Boykov et al., 2006] proposed the first known interactive graph cuts-based segmentation algorithm designed to divide an image into foreground and background regions, as displayed in Fig. 2.15. To accomplish the aforementioned objective, Boykov et al. represented an image as an undirected graph $G = (V, E)$ where a node $v_i \in V$ corresponded to a particular pixel, and ascertained two additional nodes called an object terminal (source node-s) and a background terminal (sink node-t). In this new representation of G, the set of edges E consisted of two types of distinctly weighted components called neighborhood links (n-links) that connected neighboring vertices (v_i and v_j), and terminal links (t-links) that connected pixels to source/sink nodes. The algorithm starts by allowing a user, through mouse operated brush strokes, to interactively mark a set of pixels (called 'seeds') pertaining to foreground/background content in the scene. These marked pixels were subsequently employed as hard constraints during the segmentation process. Additionally, the algorithm uses a cost function that incorporates region and boundary information imposed as soft constraints on the segmentation protocol. The aforestated cost function was predicated on specific edge weights partially derived from foreground/background intensity histograms. Finally, a unique graph cut framework based on the min-cut/max flow criterion [Boykov et al., 2001, Boykov et al., 2004] was utilized to uncover a single globally optimal solution as the final segmentation outcome, satisfying user-defined hard constraints whilst maintaining the best balance between region and boundary information. The algorithm was demonstrated on gray scale images and 3-D CT/MRI volumes.

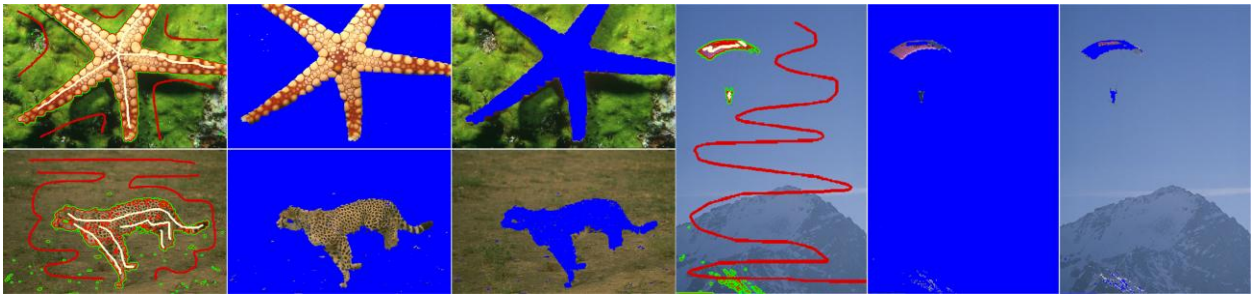

Fig. 2.15: Results of the interactive graph cuts-based segmentation algorithm in [Boykov et al., 2001] acquired using an implementation provided by [Gulshan et al., 2010]. For each image shown is user

defined foreground (white) and background (red) brush strokes as well as corresponding segmentation outcomes.

The work in [Boykov et al., 2001] was further enhanced by Rother et al. [Rother et al., 2004], who devised an interactive graph cut methodology for color imagery called 'GrabCut', where color information was incorporated into the cost function using a Gaussian mixture model. The proposed technique first acquires an optimal 'hard' partitioning result, using the aforestated graph cut algorithm enforced in an iterative scheme that fluctuates between estimation and parameter learning to solve the min-cut/max-flow criterion until convergence is achieved. Subsequently, a border matting protocol is employed as a post processing or corrective editing mechanism to arrive at the final segmentation result. Han et al. [Han et al., 2009] established a color image segmentation algorithm by extending the GrabCut methodology to accommodate color and multiscale nonlinear structure tensor texture (MSNST) features. This augmented GrabCut technique was found to have superior performance in comparison to its traditional equivalent over a diverse test bed of images taken from the Berkeley segmentation database. Li et al. [Li et al., 2004] developed an interactive foreground/background segmentation tool called 'Lazy Snapping' with instant visual feedback driven by a novel segmentation algorithm that integrated graph cuts and pre-computed over-segmentation. The proposed framework comprised of an object marking step where users indicate foreground objects of interest at a coarse scale, followed by a boundary editing step where object boundaries are delineated at a finer scale. Yang et al. [Yang et al., 2007] devised a new foreground/background segmentation algorithm called 'progressive-cut' that explicitly incorporated user evaluation, interaction, along with intention/expectation in a graph cut schema, for yielding enhanced visual feedback and segmentation accuracy with minimal interaction. Feng et al. [Feng et al., 2010] proposed an unsupervised extension of the binary (or foreground/background) s-t graph cut known as a graduated graph cut (GGC), with an architecture that possessed capabilities of self-validated labeling of MRFs. Moreover, by utilizing different MRF optimizing protocols, Feng et al. proposed three algorithms namely tree-structured graph cuts (TSGC), net structured graph cuts (NSGC), and hierarchical graph cuts (HGC) for color image segmentation.

Among other advancements concerned with graph-based segmentation, a number of approaches have been developed that fall outside the category of using 'cuts' for partitioning an image. Felzenszwalb et al. [Felzenszwalb et al., 2004] developed a tree-structured segmentation technique by defining a predicate for determining the existence of a boundary pairwise between regions, utilizing graph representations of images (see Fig. 2.16). This predicate was defined as the disparity among pixel intensities along the boundary between the regions relative to the neighboring pixel intensity differences within each of the regions. The results portrayed in Fig. 2.16, were derived using default parametric settings $sigma$=0.5, k =500 and min=20, each of which are illustrated in [Felzenszwalb et al., 2004].

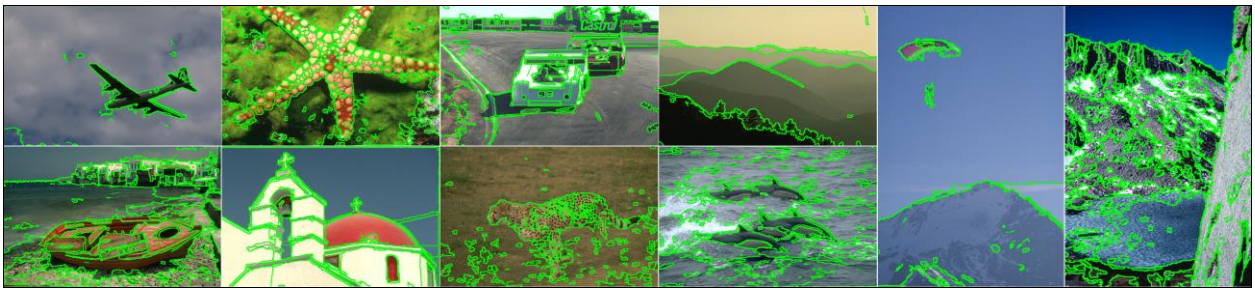

Fig. 2.16: Results of the graph-based segmentation algorithm in [Felzenszwalb et al., 2004].

Ding et al. [Ding et al., 2008] established another tree-structured graph segmentation procedure for sematic object segmentation. The work employed a scale-based connected coherence tree algorithm (CCTA) that connected all coherent/similar pixels in a graph using a specific criterion that maximized the probability of them being part of salient regions in a scene. Dupuis et al. [Dupuis et al., 2006] introduced a new color image segmentation process formulated in a graph partitioning architecture where an affinity matrix, that signified the pairwise similarity of nodes (or pixels) in a graph, was acquired using a linear combination of affinity matrices from multiple visual cues such as color, texture, gradient and luminance. The resultant affinity matrix was utilized in the partitioning process that included meta-edge computations and node contraction operations to arrive at the final segmentation result. Grady et al. [Grady et al., 2006] proposed a graph partitioning algorithm that uncovered partitions with small isoperimetric ratios, as an alternative to conventional spectral graph partitioning. The algorithm was found to be significantly faster in comparison to Ncuts-based segmentation with much more stable results. Image segmentation using random walks was first seen in the work of Grady [Grady, 2006], who presented an interactive process that assigned to every un-labeled pixel, a label that corresponded to the highest probability that a random walker originating at that pixel would first reach the assigned label, from amongst a set of user-defined ones. The algorithm was demonstrated on synthetic and real imagery with satisfactory performance. The work in [Grady et al., 2006] was extended by Yang et al. [Yang et al., 2010] using a constrained random walk-based algorithm that accommodated multiple user inputs together with local contour deformation, for enabling highly accurate and computationally efficient object segmentation in color imagery. Lately, Xiang et al. [Xiang et al., 2011] designed an interactive image segmentation algorithm using graph-based transductive classification; a procedure that involved multiple linear reconstructions in small image windows (MLRW). The algorithm was a two-step process where, in every 3x3 image window, the color intensity of the central pixel was first reconstructed using a weighted combination of its eight neighbors and the acquired optimal weights were in turn used in the second step to linearly reconstruct the class label of that pixel.

Apart from the procedures discussed in this subsection, numerous supervised/unsupervised segmentation methods involving graph-cuts [Liu et al., 2010, Ding et al., 2010, Tao et al., 2010, Salah et

al., 2011] and hybrid techniques using graph formulations [Makrogiannis et al., 2005, Makrogiannis., 2005, Sumengen et al., 2006, Wangenheim et al., 2007, Wangenheim et al., 2009, Kumar et al., 2010, Couprie et al., 2011] have been developed as tools for driving various imaging applications.

2.2.1.2.8 Watershed-based Segmentation Techniques

Over the years, methods based on morphological watersheds have carved out a niche of their own in the segmentation literature. Watershed segmentation protocols typically utilize region as well as contour information to partition an image, by viewing it as a 3-D topographic relief (see Fig. 2.17) involving two spatial dimensions and the third one being a specific attribute (eg. intensity or gradient). Conceptually, watershed algorithms identify a pixel in an image as being stationed [Gonzalez et al., 2008] in the attribute-terrain either within troughs/basins generally associated with region minimums, or at a location where a drop of water would flow towards a single or multiple region minimums. Pixel conglomerations that satisfy the first two conditions typically form catchment basins better known as watersheds; while ones that conform to the third condition (multiple region minimums) typify watershed lines. Watershed segmentation possesses several advantages such as: 1) simplicity in algorithmic design, 2) minimal computational complexity, 3) ability to provide close contours even in low contrast regions with weak boundaries, and 4) means to serve as a stable initialization for more sophisticated segmentation mechanisms. On the downside, the output achieved by a watershed transform is often over-segmented and requires post processing schemes involving region merging and markers (connected components branding flat regions or objects in images) to yield a more suitable outcome.

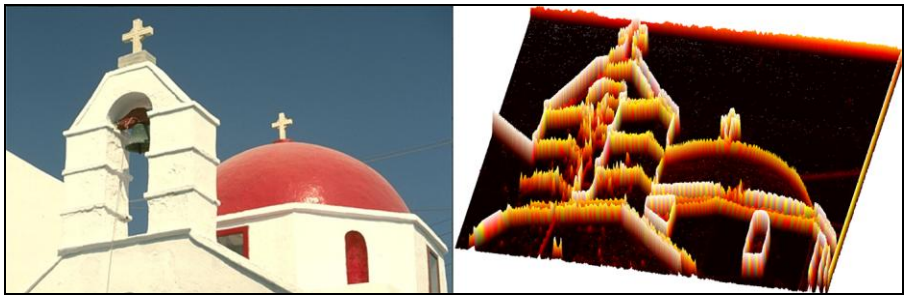

Fig. 2.17: Sample image and 3-D topographic relief of its gradient.

Gao et al. [Gao et al., 2001] developed a three-step color image segmentation approach entailing simplification, marker extraction and boundary decision processes respectively. The first step was responsible for getting rid of any unwarranted image information from a segmentation standpoint, while the second step facilitated the functionality of homogeneous region identification. Finally, a modified region-growing-based watershed algorithm was performed in the last step to determine the eventual region boundaries. In a disparate research endeavor, Gao et al. [Gao et al., 2006] devised a marker-based watershed segmentation methodology based on the concept of disjoint set union that involved pixel

sorting, set union, and pixel resolving processes. Hill et al. [Hill et al., 2003] incorporated the concept of texture gradient in a watershed segmentation framework to counter over-segmentation problems while partitioning textured regions that are perceptually homogeneous in images. The algorithm begins by extracting texture and gradient information from the scene using a non-decimated form for the complex wavelet transform. Subsequently, a new marker guided watershed algorithm was employed to identify a final set of homogeneous textured/non-textured regions. Nguyen et al. [Nguyen et al., 2003] pioneered a novel segmentation approach called "watersnakes" that unified principles from energy-based active contours with the watershed transform. Through the watersnake scheme, Nguyen et al. formulated watershed segmentation as an energy minimization problem for region/object segmentation. In a similar context of fusing energy-based methodologies and watershed segmentation, Vanhamel et al. [Vanhamel et al., 2003] utilized a vector energy-based nonlinear diffusion filtering multiscale space, where a hierarchy of strong edges, smooth segments and detailed segments were extracted by color gradient watersheds to identify meaningful regions. The aforementioned algorithm was demonstrated on simulated as well as natural scene color imagery. Kim et al. [Kim et al., 2003] proposed an effective watershed segmentation method in a multiresolution architecture using a wavelet transform. Once the image pyramid was constructed, the algorithm starts at the lowest level using a watershed segmentation step. The obtained segments were optimized using a region merging protocol and projected to the next higher resolution using an inverse wavelet transform. The aforesaid step was continued until a segmentation outcome at the highest/original image resolution was achieved. In contrast, Jung [Jung, 2007] designed a color image segmentation protocol called Waveseg using a dyadic wavelet decomposition scheme to create multiresolution representations of the input image. To this effect, a watershed transformation was applied to the image at the lowest resolution yielding an initial segmentation. This segmentation estimate was repetitively projected to subsequent higher resolutions using inverse wavelet transforms and contour refinements processes to obtain a full resolution estimate of the identified regions. Finally, a region merging protocol was exercised to merge adjoining regions with similar colors producing the output segmentation result. The algorithm was found to perform favorably in comparison to state-of-the-art methods especially when handling noisy color images. Arbeláez et al. [Arbeláez et al., 2011] introduced a three-step contemporary image segmentation strategy (abbreviated as $gPb-owt-ucm$) involving a transformation, called the Oriented Watershed Transform (OWT). In the first step, the proposed algorithm detects contours in the input image by computing a metric called the globalized probability of boundary (gPb), using brightness, color and texture cues. The identified contours (which may not be closed) are subsequently given as input to an OWT to yield an initial set of regions. Finally, an agglomerative clustering procedure is utilized to hierarchically form meaningful regions represented as a 'region-tree' by

an Ultrametric Contour Map (UCM). Fig. 2.18 illustrates the results of the $gPb-owt-ucm$ methodology achieved by thresholding the UCM at level 0.3.

In addition to the segmentation approaches discussed in this sub-section, several other hybrid methods involving the watershed transform [O'Callaghan et al., 2005, Makrogiannis et al., 2005] have been proposed, emphasizing its importance in the segmentation realm.

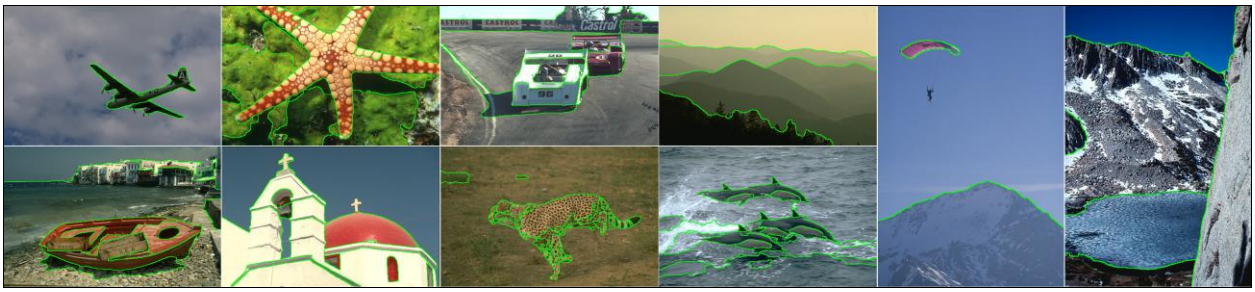

Fig. 2.18: Results of the $gPb-owt-ucm$ algorithm [Arbeláez et al., 2011].

As mentioned at the beginning of Section 2.2.1, a major portion of segmentation practices can be viewed as being either spatially blind or spatially guided, and within the notion of each of these principles, dominant groups of methods have been discussed. However, there are several techniques that may not distinctly fall in any of the aforestated dominant categories but nonetheless provide valuable contributions to the field of image segmentation. Consequently, we will categorize them as a separate group called 'miscellaneous' approaches. Among these are: 1) fuzzy-based procedures such as the ones involving fuzzy homogeneity [Cheng et al., 2003, Chaabane et al., 2010] and fuzzy region completion [Choy et al., 2011], 2) supervised techniques using adaptive weighted distances [Protiere et al., 2007], spline regression [Xiang et al., 2009], geodesic matting [Bai et al., 2009] and linear programming [Li et al., 2010], 3) methods using specialized image features namely Quaternions [Funt et al., 2007, Subakan et al., 2011], textons [Shotton et al., 2006, Shotton et al., 2007, Shotton et al., 2008, Yu et al., 2012], Histogram of Oriented Gradients (HOG) [Wang et al., 2008], and Local Binary Patterns (LBP) [Nammalwar et al., 2010, Li et al., 2011], 4) methodologies that employ turbo-pixel or super-pixel based representations of image data [Rohkohl et al., 2007, Levinshtein et al., 2009, Achanta et al., 2010, Liu et al., 2011, Artan, 2011, Li et al., 2012], 4) physics-based processes robust to shadowing, shading and highlighting effects [Hoang et al., 2005 (see Fig. 2.19 for sample segmentation outcomes with default parameters), Vazquez et al., 2011], 5) routines that treat segmentation as a classification task using sophisticated classifiers such as Support Vector Machines (SVMs) [Wang et al., 2011, Yu et al., 2011] or employ specialized properties of images [Macaire et al., 2006], 6) top down (TD) or bottom up (BU) schemes using shape constraints [Zöller et al., 2007], as well as integrated TD-BU frameworks [Borenstein et al., 2008], 7) mechanisms involving statistical principles [Delyon et al., 2006], information bottleneck method [Bardera et al., 2009] and algorithms that consider segmentation as a task of finding

perceptually salient groupings [Song et al., 2011, Usó et al., 2011], and finally 8) co-clustering strategies which combine multiple segmentations into one improved result [Ghosh et al., 2009, Vitaladevuni et al., 2010, Glasner et al., 2011] as well as co-segmentation methods that jointly segment multiple images which contain a common object [Rother et al, 2006, Cheng et al., 2007, Hochbaum et al., 2009, Mukherjee et al., 2009, Batra et al., 2010, Chang et al., 2011, Vicente et al., 2011, Joulin et al., 2012, Kim et al., 2012]. Co-clustering and co-segmentation (referenced above) are amongst the newest techniques being researched in the segmentation domain. Co-clustering is formally defined as the process of jointly clustering two or more images that closely maintain their semantic foreground and background content such as shapes, color, and texture of objects/regions. Examples of such a set of images include digital video frames in close proximity, series of images taken under varying camera exposure and/or illumination conditions. On the other hand, co-segmentation is a procedure wherein multiple images that have diverse backgrounds are processed to segment common foreground objects well-correlated in terms of their color-texture composition. Co-segmentation has several applications such as image editing, image similarity measurement, video summarization/tracking, and object-based image retrieval to name a few.

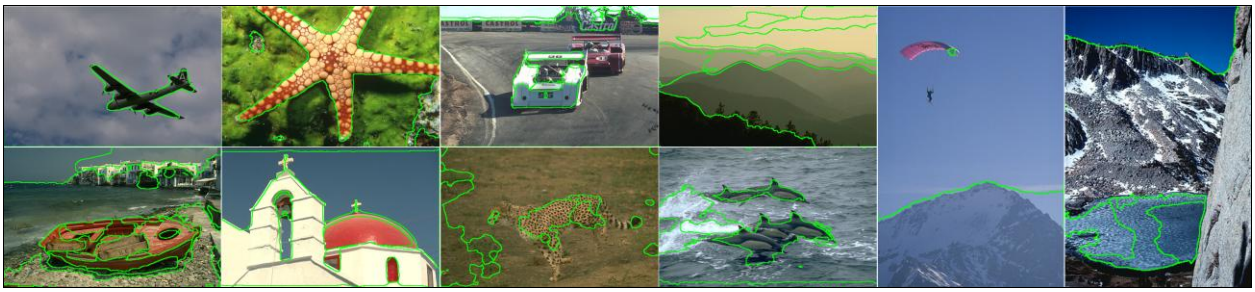

Fig. 2.19: Results of the algorithm in [Hoang et al., 2005].

2.2.2 Multi/Hyperspectral Image Segmentation Methodologies

Similar to the color/gray scale image segmentation domain, in the recent past, several approaches [Carleer et al. 2005] have been developed for multi/hyperspectral image segmentation which can be broadly categorized as being either: 1) spatially blind or 2) spatially guided.

Spatially blind techniques such as clustering [Xu and Wunsch, 2005] and thresholding typically segregate pixels in a specific feature space, and for the most part of their modus operandi disregard spatial relationships among pixels. Consequently, regions derived out of such procedures may or may not be spatially compact. Clustering, a method which has had wide spread use for segmentation of remotely sensed data, generally analyzes an image perceived as a point cloud in a spectral space and partitions it using objective functions to identify pixel groupings. The aforementioned partitioning is done such that when concluded, pixels within a cluster have a high degree of conformity while cross cluster groups have low compatibility. Bilgin et al. [Bilgin et al., 2011] proposed a segmentation routine for hyperspectral imagery utilizing a new subtractive clustering and validation methodology involving One Class Support

Vector Machines (OC-SVM). Mercovich et al. [Mercovich et al., 2011] instituted a clustering technique for multispectral images based on the principle of optimal modularity that finds communities within the data viewed as a graph. Among hybrid practices, Tran et al. [Tran et al., 2003] developed a Spatial Refinement (SpaRef) clustering approach for multispectral images by integrating spatial information into a partitioning framework involving the K-means and agglomerative hierarchical clustering protocols. In their work, Lee and Lee [Lee and Lee, 2010] introduced a segmentation routine for hyperspectral data using mean shift clustering. The algorithm is initiated by a Principle Component Transformation (PCT) of the input image. Subsequently, the first three components are fed as input to a mean shift filtering process. Finally, the resultant over-segmented clusters are merged together in a region merging stage using the Bhattacharya distance.

Contrary to aforestated mechanisms, spatially guided or image domain-based schemes take into account spatial affiliations among pixels during the segmentation process and consequently tend to generate compact regions. Methods such as growing, splitting, merging, morphological operations, watersheds as well as energy driven schemes have been effectively used for segmenting remote sensing data. Paglieroni [Paglieroni, 2003] devised a multi-band region growing approach explicitly based on spectral information for segmenting hyperspectral images by generalizing a standard region growing method. On the other hand, Lee and Lee [Lee and Lee, 2008] constructed a methodology for unsupervised segmentation of hyperspectral data based on Canny edge detection, region growing and morphological region filling procedures. Gorretta et al. [Gorretta et al., 2009], introduced a butterfly approach for hyperspectral image segmentation in an iterative framework comprising of a cross analysis of spectral-spatial information, using a split and merge set up and the normalized cuts algorithm. Cartographic data guided segmentation of very high spatial resolution multispectral imagery using a region growing strategy was first seen in the work of Bouziani et al. [Bouziani et al., 2010]. This technique was novel in the sense that information derived from pre-existent digital maps of a scene was leveraged to provide effective segmentations. Akçay and Aksoy [Akçay and Aksoy, 2008] proposed an algorithm for segmenting hyperspectral image information through region profiles derived from morphological opening and closing operations, applied to individual spectral bands using structuring elements of increasing sizes. The underlying idea of their segmentation process was to combine spectral information from the input data cube with structural information extracted through the obtained morphological profiles.

Within notion of watershed-based processing, Li and Xiao [Li and Xiao, 2007] developed a segmentation framework by extending the classical watershed transformation for multispectral images. More recently, Li et al. [Li et al., 2011] integrated the work in [Li and Xiao, 2007] with the dynamics of watershed contours to develop a hierarchical segmentation approach for very high resolution multispectral imagery. In other advances, Tarabalka et al. [Tarabalka et al., 2010] developed a segmentation framework

by extending the watershed transformation for hyperspectral images, while Angulo and Velasco-Forero [Angulo and Velasco-Forero., 2010] incorporated a semi-supervised stochastic watershed-based algorithm for hyperspectral image segmentation. Furthermore, Castilla et al. [Castilla et al., 2008] designed a segmentation method for remotely sensed data in a split-and-merge architecture that employed watershed-based partitioning and region merging. The split phase was accomplished through gradient watersheds to form an initial set of regions. The resultant set of regions was refined in a merging phase that employed size constraints and spectral distance criterions to combine various regions.

A popular category of segmentation methodologies in the image-domain realm are ones that are driven by energy formulations. Among these, region energy formulations involving Bayesian principles such as the Maximum A-posteriori Probability (MAP) criterion integrated with random field models (like Markov Random Fields (MRFs), Conditional Random Fields (CRFs)) have been extensively used for multi/hyperspectral image segmentation. Rand and Keenan [Rand and Keenan., 2001] pioneered a segmentation protocol for partitioning hyperspectral imagery using a Gibbs Random Field (GRF) conceptualization of the underlying region process (an approximation to an MRF model under certain constraints), targeted at terrain applications. A unified framework for MAP estimation applied to remote sensing image segmentation was first seen in Farag et al. [Farag et al., 2005], where the class conditional probability estimates were obtained using a supervised Support Vector Machine (SVM) density estimation method, while the class prior probability was modeled using an MRF framework that employed an analytical algorithm to automatically identify its parameters. Li et al. [Li et al., 2010] established a novel semi-supervised algorithm for hyperspectral image segmentation using multinomial logistic regression (MLR) with active learning, in an MRF formulation. The work in [Li et al., 2010] was generalized to an unsupervised algorithm in [Li et al., 2011], while the MLR procedure was integrated with a subspace projection method for superior performance in [Li et al., 2011]. In a similar context, Borges et al. [Borges et al., 2011] introduced an automatic approach to segment hyperspectral imagery based on MLR with discriminative class learning. In their work, Sun et al. [Sun et al., 2011] demonstrated a segmentation algorithm for hyperspectral data using a spatially and spectrally constrained CRF model. In addition to the spatially-guided segmentation mechanisms discussed in this section, several other image-domain methods using Partial Differential Equations (PDEs) [Ball and Bruce, 2007, Duarte-Carvajalino et al., 2008], active SVMs [Mitra et al., 2004], probabilistic neural networks [Gang and Xingjian, 2008], region-markers [Tarabalka et al., 2010] and Trilateral filtering [Wun and Messinger, 2011] have been developed.

Multi/hyperspectral image segmentation using textural features has been the crux of several research endeavors. Hazel [Hazel, 2000] proposed a methodology for multispectral scene segmentation using Gaussian Markov Random Field (GMRF) texture models, tested on data acquired from the Daedalus

sensor. Hong et al. [Hong et al., 2003] instituted an approach for segmentation of hyperspectral imagery using textural features derived from an octave-band bidirectional filter bank. Gaetano et al. [Gaetano et al., 2009] established a procedure utilizing textural features in a split-and-merge framework for the hierarchical segmentation of remote sensing images. Li et al. [Li et al., 2010] designed a unique texture-preceded segmentation algorithm applicable to high-resolution remote sensing images involving texture clustering for initial region formation followed by a fast region merging protocol based on graph models, to obtain the final segmentation outcome. Despite the numerous strides made in the field of multi/hyperspectral image segmentation, there has been relatively less development of algorithms that efficiently leverage and simultaneously fuse spectral intensity, gradient and textural information for its use in various remote sensing applications.

2.2.3 3-D Segmentation Methodologies for Medical Volumes

Many automated/semi-automated methodologies [Pham et al., 2000, Bulu and Alpkocak, 2007, Withey and Koles, 2007] have been developed to confront the segmentation problem for Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) volumes in 3-D. Approaches involving surface/contour detection clustering, thresholding, region growing, watershed-based, active contours/surfaces, texture-based, multiresolution statistics, as well as model/atlas based methods have had wide spread use for volumetric segmentation of medical data.

Early approaches developed for 3D segmentation in the medical realm, were mostly targeted at MRI volumes. Bomans et al. [Bomans et al., 1990] proposed an approach for segmentation of MRI data based on a 3-D extension of the Marr-Hildreth edge detection scheme. The fundamental principle of their approach was that the zero crossings of the aforesaid edge detector in 3-D have close correspondences to the anatomical surfaces inherent in data. In their work, Yan and Karp [Yan and Karp, 1994] devised a method based on the work in [Pappas, 1992] for segmentation of MRI volumes of the brain into K disparate tissue types using an adaptive K-means clustering algorithm, where-in each tissue was modeled by a GRF. Furthermore, Choi et al. [Choi et al., 1997] used a 3D MRF model for volumetric object extraction, demonstrated on MRI's volumes of the knee joint. Justice et al. [Justice et al., 1997], developed a semi-automated Seed Region Growing (SRG) approach, demonstrated for MR brain segmentation. Though the method is simple and easy to implement it suffers from the drawback that all volumetric elements are processed sequentially using a specific sorted order. Sijbers et al. [Sijbers et al., 1997], propounded a two-step procedure for segmentation of MR data. Initially, the entire volume was segmented using the watershed transform followed by a volume merging process based on the Minimum Description Length principle. Hastreiter and Ertl [Hastreiter and Ertl, 1998] exploited the concept of

"Intelligent Scissors" which enables propagation of contours detected in 2D to adjacent slices in 3D, to come with a methodology for brain segmentation in MRI volumes.

Over the past decade, the advancement of CT and ultrasound technology for medical imaging led to the development of several 3D segmentation algorithms for these modalities, while being compatible with MRI data. Yezzi et al., [Yezzi et al., 1997] came up with an image feature metrics driven parametric active contour model for segmentation MRI, CT and ultrasound imagery. The work was novel in the sense that it unified geometric curve evolution formalisms with traditional parametric energy models to yield the final segmentation outcome. Baillard et al. [Baillard et al., 2000] designed an energy-based geometric active surface evolution model using a Level Set framework, for robust segmentation of anatomical structures in volumetric MRI and ultrasound data. Two major advantages of their algorithm was its potential for segmenting complex anatomical structures without extensive a-priori information, and its generic nature that required minimal parametric tuning. Tek et al. [Tek et al., 2002] proposed a method for segmentation of 3D structures in CT and MRI volumes based on ray propagation by mean shift analysis that incorporates a smoothness constraint. Freedman et al. [Freedman et al., 2005] founded a model-based segmentation algorithm grounded on probability distributions of photometric variables for deformable objects in CT volumes. On the other hand, Ding et al. [Ding et al., 2005] established a single 2D atlas-guided segmentation algorithm for 3D segmentation of CT volumes, to overcome problems of complexity and control germane to 3D atlases. A prominent semi-automated method for efficient segmentation of anatomical structures was proposed by Yushkevich et al. [Yushkevich et al., 2006] based on the notions of geodesic active contours and region completion. Shen et al. Recently, [Shen et al., 2009] proposed a 3D segmentation procedure applicable to CT and MRI modalities using a novel volumetric deformable model also known as Active Volume Model (AVM), which is a self-constrained generative object model that does not require any manual training and focuses on accurate modeling of the attributes associated with the volumetric object of interest. The work in [Shen et al., 2009] was enhanced by Shen et al. in [Shen et al., 2009], where multiple-surface AVM was proposed to simultaneously segment coupled medical objects. Olowoyeye et al. [Olowoyeye et al., 2009] pioneered a medical (CT and MRI) volume segmentation methodology using its textural properties derived via a 3-D Gabor filter bank whose number depended on the size of the volume. The algorithm commences by convolving the input volume with the aforestated filter bank. Subsequently, the obtained filter responses were clustered to arrive at the final segmentation output. Campadelli et al. [Campadelli et al., 2010] developed a gray-level based segmentation framework for abdominal organs from CT scans based on a multiplanar fast marching method. Medical volume segmentation using multiresolution analysis via the 3-D discrete wavelet and ridgelet transforms integrated with Hidden Markov Models (HMMs) was first seen in the work of AlZu'bi and Amira [AlZu'bi and Amira, 2010]. In a similar context using HMMs, Huang et al. [Huang et

al., 2010] established a 3D segmentation algorithm utilizing a rotationally invariant 3D region-based Hidden Markov Model (rbHMM).

Besides the aforementioned approaches, several segmentation methods for specific abdominal components (organs as well as vertebral structure) have been developed for CT volumes. Hu et al. [Hu et al., 2001] developed a method for accurate segmentation of the lungs from CT imagery using optimal gray level thresholding and morphological operations. Kang et al. [Kang et al., 2003] instituted a multistep approach for precise segmentation of skeletal structures, using 3-D region growing and boundary refinement. Ecabert et al. [Ecabert et al., 2008] established an effective model-based approach for 3-D segmentation of the heart from CT images, using the Generalized Hough Transform. In their work, Isgum et al. [Iˇsgum et al., 2009] utilized a multi-atlas based contemporary approach for aortic and cardiac segmentation in CT volumes.

2.3. Challenges with Conventional Techniques

This chapter presents an extensive review of state-of the-art color, multi/hyperspectral and biomedical image segmentation methodologies. Despite the existence of a large number of approaches there are multiple shortcomings that need to be addressed. Several approaches generally: 1) handle only one or a few regions at a time, 2) perform pixel by pixel processing/agglomeration for region formation, 3) need manual selection of starting points for region formation, 4) have processing order dependency (rasterized architecture), 5) lack the feasibility for practical usage because of intense computations. Furthermore, there are not many segmentation algorithms that have been developed with a multi-modal perspective, to simultaneously cater to the requirements of sophisticated applications that employ data sets from remote sensing and biomedical modalities whilst being backward compatible with applications using gray scale or color imagery. We propose a segmentation framework that identifies meaningful constituents in image data using multiple image features, tested across three target imaging modalities (conventional color, remotely sensed, and bio-medical volumetric imagery) and in doing so overcoming the aforestated shortcomings.

Chapter 3

Automatic Color Image Segmentation by Dynamic Region Growth and Multivariate Region Merging

An overview of the proposed Gradient-SEGmentation (GSEG) algorithm, primarily consisting of four modules is shown in Fig. 3.1. The first module (M1) performs the functionalities of color space conversion of the input image, vector gradient computation, and threshold calculations for subsequent processing. The second module (M2) performs region growth and dynamic seed addition which dynamically select regions of contiguous pixels that display similar gradient and color values, producing an initial segmentation map. The third module (M3) creates a texture characterization channel by first quantizing the input image, followed by entropy based filtering of the quantized colors of the image. Finally, the GSEG algorithm culminates in the fourth module (M4) which performs a region merging procedure utilizing the results of the prior two modules. The following sub-sections describe each of the three modules in detail.

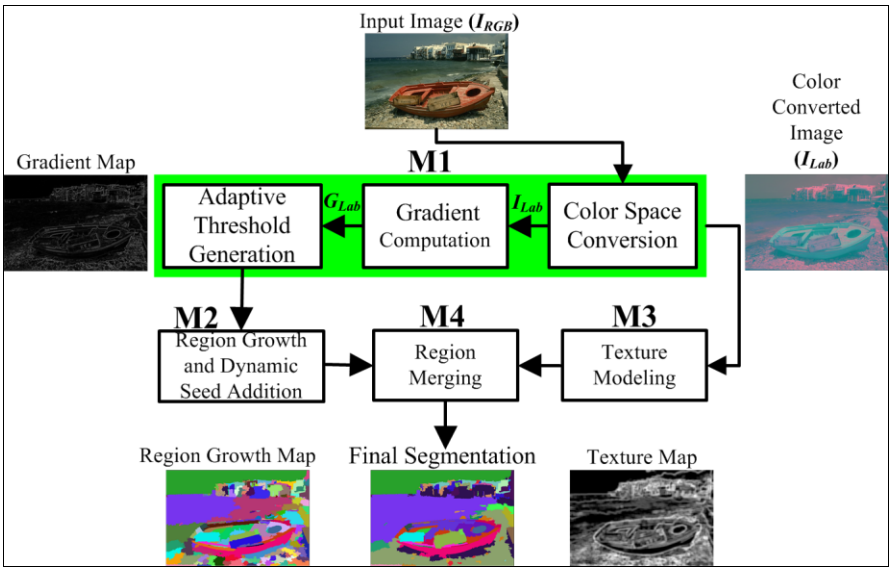

Fig. 3.1: Block diagram of the proposed Gradient SEGmentation algorithm

3.1 Color Space Conversion

The GSEG algorithm begins with a conversion of the input image (I_{RGB} of dimensions d_M by d_N) from RGB to CIE L*a*b (I_{Lab}) for improved correct color differentiation, owed to the fact that the latter is better modeled for human perception in comparison to the RGB space. This is justified by the fact that given two colors, the magnitude difference of the numerical values between them is proportional to the

perceived difference as seen by the human eye [Green and MacDonald, 2002], a property that cannot be associated with the RGB space. The L*a*b* data is 8-bit encoded to values ranging from 0-255 for convenient color interpretation, and to overcome viewing/display limitations. The resultant color converted data is subsequently employed for computing the vector color gradient (G_{Lab}) utilizing the algorithm described in section 3.2. In general for an arbitrary image, the gradient values for 8-bit L*a*b* data were found to span over a much smaller range than 8-bit RGB, consequently resulting in a relatively compact histogram than its RGB counterpart. In Figs. 3.2(a) and 3.2(b), shown are the plots of gradient information, computed in the RGB and L*a*b* color spaces, of the 'Cars' and 'Cheetah' images respectively. From both these images it can be observed that the gradient values computed in the L*, a*, b* color space span a significantly smaller range, than the RGB color space. This is due to the fact the range of luminance and chrominance values are significantly lower than red, green and blue values. Furthermore, since the GSEG algorithm was designed such that it is iterative and dynamic over the gradient range of the input image, the use of the CIE L*a*b* color space reduced the computational requirement for region processing as the algorithm is confined to a significantly smaller range of gradient values, as opposed to its RGB counterpart.

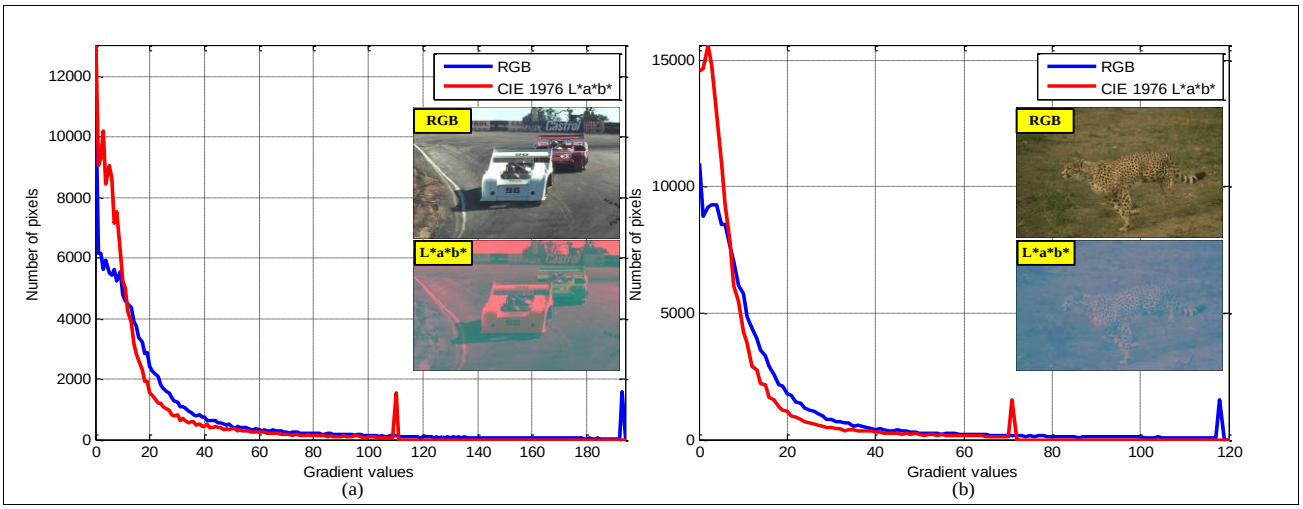

Fig. 3.2: Comparison of gradient information (RGB vs. CIE L*a*b*) of: (a) Cars, (b) Cheetah.

3.2 Edge detection for Color Imagery

The proposed algorithm uses an edge-detection algorithm [Lee and Cok, 1991] that provides the magnitude of edges present in an image. These edges help to detect the individual regions into which an image is segmented and the direction in which the region growth procedure takes place. The detected areas with no edges inside them are identified as the initial clusters or seeds selected to initiate the segmentation process.

In what follows, let $I_{3\to 2} = [I_1, I_2, I_3]^T$ be a two-dimensional (spatial) vector field representing a color image with three attributes (such as red, green and blue or L*, a*, b* channels), each of dimensions d_M x d_N. Furthermore, let x, y denote the spatial co-ordinates in an image band such that $1 \le x \le d_M$, and $1 \le y \le d_N$. Consequently, a spatial gradient matrix (J) of all first order partial derivatives (also known as the Jacobian matrix), at pixel location (x, y) can be defined as:

$$J(x,y) = \begin{bmatrix} \frac{\partial I_1(x,y)}{dx} & \frac{\partial I_1(x,y)}{dy} \\ \frac{\partial I_2(x,y)}{dx} & \frac{\partial I_2(x,y)}{dy} \\ \frac{\partial I_3(x,y)}{dx} & \frac{\partial I_3(x,y)}{dy} \end{bmatrix} \tag{3.1}$$

The magnitude of the gradient at location (x, y) is acquired by first constructing an inner product matrix $J^T J(x,y)$, given by:

$$J^T J(x,y) = \begin{bmatrix} a(x,y) & c(x,y) \\ c(x,y) & b(x,y) \end{bmatrix} \tag{3.2}$$

where,

$$a(x,y) = \left(\frac{\partial I_1(x,y)}{dx}\right)^2 + \left(\frac{\partial I_2(x,y)}{dx}\right)^2 + \left(\frac{\partial I_3(x,y)}{dx}\right)^2 \tag{3.3}$$

$$b(x,y) = \left(\frac{\partial I_1(x,y)}{dy}\right)^2 + \left(\frac{\partial I_2(x,y)}{dy}\right)^2 + \left(\frac{\partial I_3(x,y)}{dy}\right)^2 \tag{3.4}$$

$$c(x,y) = \left(\frac{\partial I_1(x,y)}{dx}\right)\left(\frac{\partial I_1(x,y)}{dy}\right) + \left(\frac{\partial I_2(x,y)}{dx}\right)\left(\frac{\partial I_2(x,y)}{dy}\right) + \left(\frac{\partial I_3(x,y)}{dx}\right)\left(\frac{\partial I_3(x,y)}{dy}\right) \tag{3.5}$$

Following this, the gradient magnitude (G) is computed by finding the square root of the largest Eigenvalue of $J^T J(x,y)$, given by:

$$G(x,y) = \sqrt{\arg\max_i \{\lambda_i [J^T J(x,y)]\}} \tag{3.6}$$

In Eq. (3.6) $\lambda_i [J^T J(x,y)]$ is the i^{th} Eigenvalue of $J^T J(x,y)$. Additionally, it can be recognized that that $J^T J(x,y)$ has two Eigenvalues ($\lambda_i, i = 1, 2$) calculated specifically at the pixel site (x, y). Also, since $J^T J(x,y)$ is a real, symmetric and positive semi-definite 2x2 matrix for every pixel location, the close form solution of its largest Eigenvalue can be given as:

$$\arg\max_i \{\lambda_i [J^T J(x,y)]\} = \frac{1}{2}\left\{a(x,y) + b(x,y) + \sqrt{(a(x,y) + b(x,y))^2 - 4(a(x,y)b(x,y) - c(x,y)^2)}\right\} \tag{3.7}$$

3.3 Adaptive Threshold Generation

Having obtained the vector color gradient in the CIE L*a*b* color space, the next stage of the GSEG involves an adaptive threshold generation scheme, to calculate thresholds required for region formation and processing. Region formation is performed in two distinct phases: 1) the first phase called the Initial Clustering phase, involves identifying a few potential homogeneous sites in the image where seeds (pixel clusters that are assigned labels corresponding to various image regions) are placed, 2) the next phase of region processing called as Region Growth and Dynamic Seed Addition (RGDSA) phase, comprises the growth of the initially places seeds until they reach their maximum size, accompanied by the addition of new seeds in regions that remain unsegmented at specific stages of processing. The two phases are carried out by thresholding the histogram of the gradient map, commencing in low gradient areas called 'color homogeneous regions' and culminating in strong gradient areas called 'regions of color transitions'. The primary objective of the adaptive threshold generation scheme is to provide suitable thresholds to delineate the instants at which seed addition should be performed during region growth (in a procedure called 'dynamic seed addition' described later).

A threshold value (T_n) constraining an arbitrary n^{th} stage of region growth, simultaneously administering the addition of new seeds in unprocessed image regions, is evaluated as a function of the cumulative areas of the gradient histogram greater and lesser than the limiting threshold (T_{n-1}) value of its corresponding previous stage ($(n-1)^{th}$ stage), as illustrated in Fig. 3.3. This is done in order to: 1) make certain that regions of significant size are always added to the segmentation map, 2) ensure that all thresholds lie within span of the histogram, avoiding the possibility of wasted computational efficiency, and 3) account for the exponential decay of the gradient map histograms of natural scene images (as seen in Fig. 3.2). Thus if the vector color gradient map of an arbitrary image has G distinct gradient levels, and N_g is the total number of pixels that possess a gradient value g (ranging from 0 to $G-1$), then the threshold value (T_n) for the n^{th} stage of the dynamic seed addition procedure is determined by:

$$T_n = \sum_{g=0}^{T_{n-1}} N_g + n\Delta g[\sum_{g=T_{n-1}+1}^{G-1} N_g] \tag{3.8}$$

The first summation in Eq. (3.8) represents the cumulative image area less than the gradient threshold T_{n-1} that is processed in the $(n-1)^{th}$ stage of region growth, while the second summation represents the cumulative unprocessed image area greater than T_{n-1} (see Fig. 3.3). The quantity $n\Delta g$ defined as the 'growth factor', determines the incremental percentage of image area of higher gradient densities to be processed in the n^{th} stage. The entire quantity beyond the '+' sign, is known as a Region Growth Interval (*RGI*), which represents the range of gradient values from T_{n-1} to T_n (lower and upper limits of the n^{th} *RGI*), as depicted in Fig. 3.3. In this manner utilizing Eq. (3.8) and an initialization threshold (T_0: employed in the Initial Clustering phase, discussed in the next section), segmentation thresholds T_1 to T_n

differentiating *'n' RGIs* are computed, which are utilized for the previously mentioned functionalities. It's critical to recognize that the aforementioned thresholds are selected in the gradient range greater than the initialization threshold, consequently ensuring that all regions of higher gradient densities are processed.

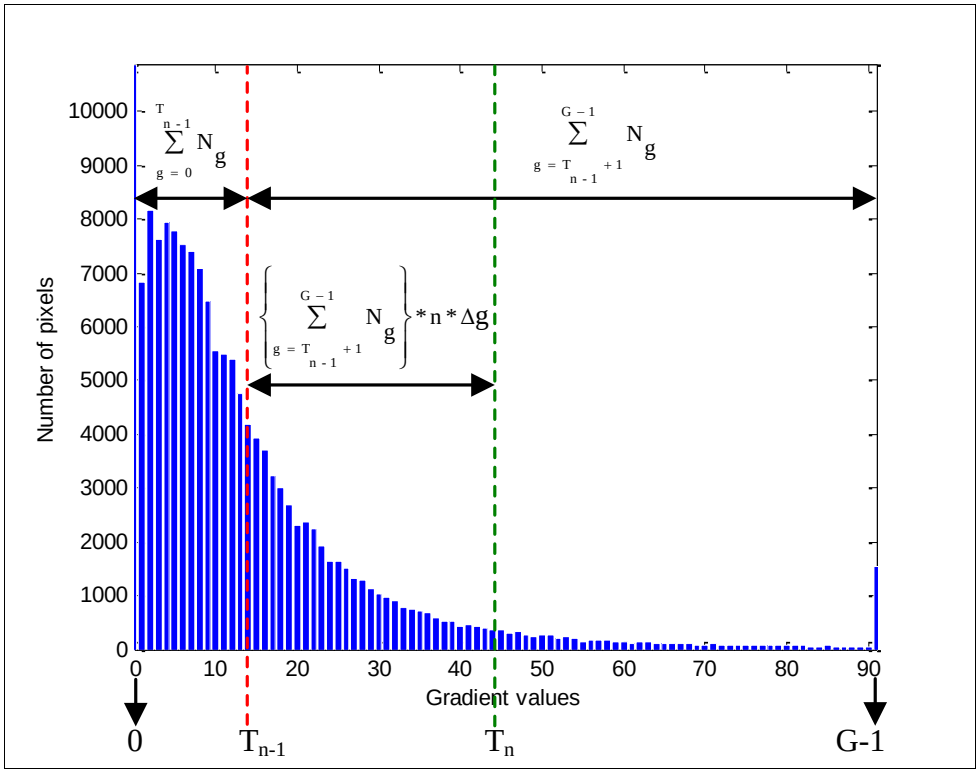

Fig. 3.3: Adaptive threshold generation scheme for Region Growth and Dynamic Seed Addition.

3.4 Initial Clustering

As mentioned previously the initial objective is to identify potential homogeneous sites in the image, where seeds can be placed (see Fig 3.4). Furthermore since the goal of the GSEG is to perform segmentation beginning in color homogenous regions corresponding to low gradient region in the image, we propose identifying homogeneous seed sites by choosing a threshold value in extremely low gradient ranges. This initialization threshold (T_0 =5) was empirically chosen based on extensive testing over several hundred images. All the pixels in the gradient map that fall below this initialization threshold are considered to be pixels in 'flat' regions, which are subjected to connected component analysis to generate the initial seed map. This initial seed map is also called as the *Parent Seeds (PSs)* map, as shown in Fig. 3.4 (d). The labeling procedure uses the general procedure outlined in reference [Haralick and Shapiro, 1992]. It is briefly described as follows: 1) run-length encoding of the input image, 2) scan the runs, and assign preliminary labels and recording label equivalences in a local equivalence table, 3) resolve the equivalence classes, and 4) re-label the runs based on the resolved equivalence classes. Moreover to prevent the generation multiple clusters within homogeneous and connected regions, parent seeds are

constrained to clusters of pixels which are greater than certain size criterion Fig. 3.4 (e). All size criterions in the GSEG algorithm are a function (varying multiplicative products) of a Minimum Seed Size (MSS) value, defined as the smallest possible size of a region desired in the output segmentation.

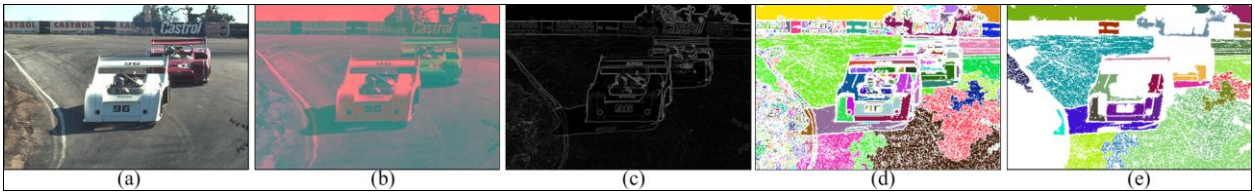

Fig. 3.4: (a) Input image I_{RGB}, (b) color converted image) I_{Lab}, (c) Gradient map G_{Lab}, (d) parent seeds (*PSs*) map, (e) final *PSs* map after employing size constraints.

The *MSS* criterion for region processing is chosen such that it is very small in comparison to the corresponding image area (that is $MSS << \{d_M d_N\}$) to ensure that fine details are captured in the segmentation process. Consequently, based on the aforementioned requirements the *MSS* criterion is computed as:

$$MSS = \alpha\{d_M d_N\} \tag{3.9}$$

where α is a small percentage (typically 0.01%) of the input image area. If the initialization threshold is unsuccessful in generating parent seeds, the threshold value is automatically increased until at least one region is detected. Moreover, the size criterion for addition of new seeds at subsequent higher gradient densities is gradually reduced for proper region formation. The generation of parent seeds concludes the Initial Clustering phase.

3.5 Region Growth and Dynamic Seed Addition (RGDSA)

The quality of region-growth techniques is highly dependent on the locations chosen to initialize the growth procedure. We propose an alternate process for region growth that does not depend exclusively on the initial assignment of clusters for the final segmentation. The procedure searches for regions where the gradient map displays no edges. The selected regions form the initial set of seeds to segment the image. The region growth procedure also accounts for regions, which display similar edge values throughout, by detecting unattached regions at various edge density levels. Fig. 3.5 displays a flowchart of this module, followed by a detailed explanation.

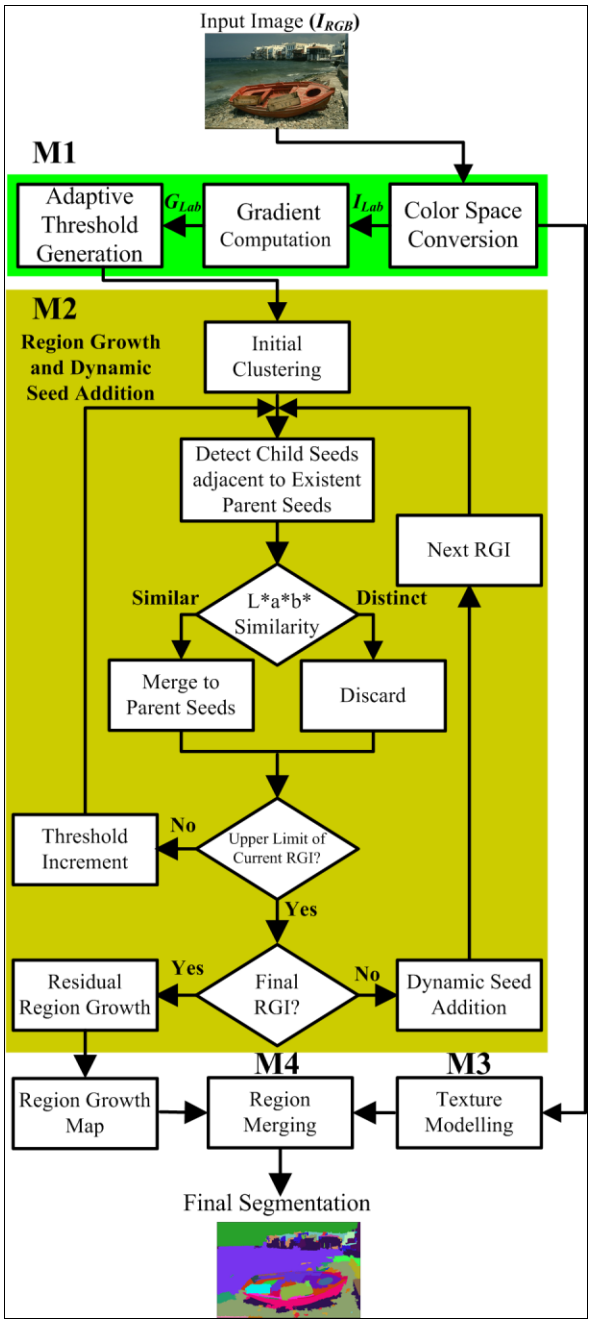

Fig. 3.5: Flowchart of the region growth and dynamic seed addition process

Once the parent seeds are generated (see Fig. 3.4(d) or Fig. 3.6(d)), the GSEG algorithm proceeds to the region growth procedure performed over all Region Growth Intervals (*RGIs*). In an indiscriminate *RGI* spanning from T_{n-1} to T_n as seen in Fig. 3.3, where $1 \le n \le N$ for N-stages of growth, the proposed algorithm proceeds by a sequential increment of the threshold from T_{n-1} to $T_{n-1}+1$ to detect new areas with higher gradient densities. All the pixels in the gradient map whose value lies in between T_{n-1} to $T_{n-1}+1$ undergo connected component labeling, to detect new areas with higher gradient densities, referred to as

Child Seeds (*CSs*), as shown in Fig. 3.6(e). However, only the *CSs* that are adjacent to existent *PSs* are administered at this point. Adjacent *CSs* (Fig. 3.6 (g)) are found by obtaining the ones that share pixels with the boundaries of *PSs* (depicted in Fig. 3.6(f)).

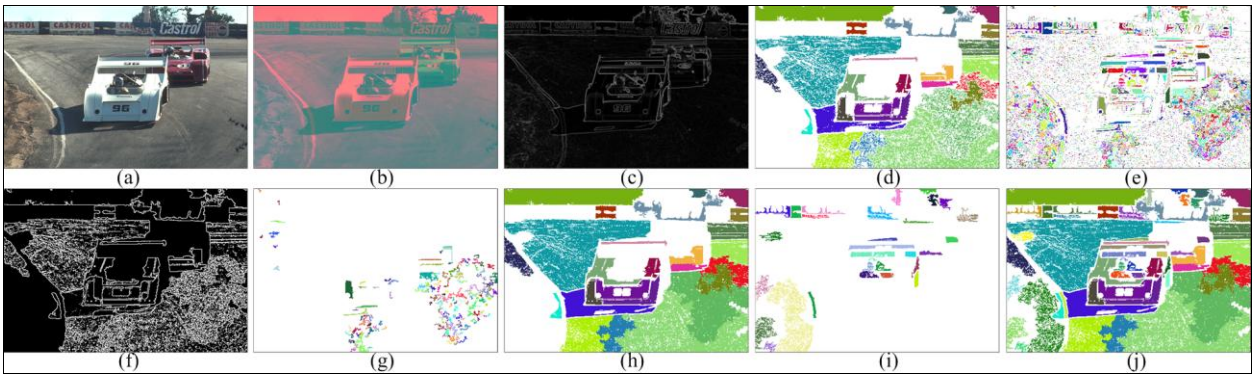

Fig. 3.6: (a) Input image I_{RGB}, (b) color converted image I_{Lab}, (c) Gradient map G_{Lab}, (d) final *PSs* map after employing size constraints, (e) child seeds (*CSs*) generated on threshold increment, (f) parent seed boundaries $PS_{boundaries}$, (g) adjacent child seeds, (h) seed map at the end of 1st region growth interval, (i) spatially independent seeds generated during dynamic seed addition, (j) parent seed map for next stage of region growth.

Parent Seed boundaries ($PS_{boundaries}$) is a binary map extracted morphologically by subtracting *PSs* from their corresponding dilated counterparts, mathematically represented as:

$$PS(i,j)_{boundaries} = \begin{cases} 1 & \forall\ (i,j)\ where\ \{(PSs \oplus \psi) - PSs\} > 0 \\ 0 & \forall\ (i,j)\ where\ \{(PSs \oplus \psi) - PSs\} = 0 \end{cases} \tag{3.10}$$

$$PSs \oplus \psi = \{\gamma \in Z^2 \mid \gamma = a + b \quad for \quad some \quad a \in PSs, b \in \psi\} \tag{3.11}$$

where *(i,j)* is a random pixel co-ordinate and $PSs \oplus \psi$ represents the dilation of the *PSs* map with a 3X3 square structuring element ψ that possesses a value of '1' at every location. Following this, to facilitate minimal computational expense during region growth, the Minimum Seed Size (*MSS*) criterion is utilized to differentiate between *CSs* that can directly be merged with corresponding parents and those that have to be subjected to further processing. To this effect, all the *CSs* that are smaller than the *MSS* are merged with their corresponding adjacent parents, while the ones larger than the *MSS* criterion are checked for the similarity of their lightness, red-green and blue-yellow values with surrounding *PSs*, to ascertain homogeneous segmentation. The child seed sizes are computed utilizing sparse matrix storage techniques to allow for the creation of large matrices with low memory costs. Sparse matrices store only the nonzero elements of a matrix, together with their location in the sparse matrix (indices). The size of each child seed is computed by creating a matrix of $d_M * d_N$ columns by *C* rows, where d_M is the number of columns of pixels in the image, d_N the number of rows, and *C* the number of adjacent child seeds. The matrix is

created by allocating a 1 at each column in the row that matches a pixel label. The pixels that do not have a label are ignored. To this effect, by summing all the elements along each row, we obtain the number of pixels per child seed. This procedure (see Fig. 3.7), is useful for any operation that requires the knowledge of the number of elements per group in the segmentation algorithm.

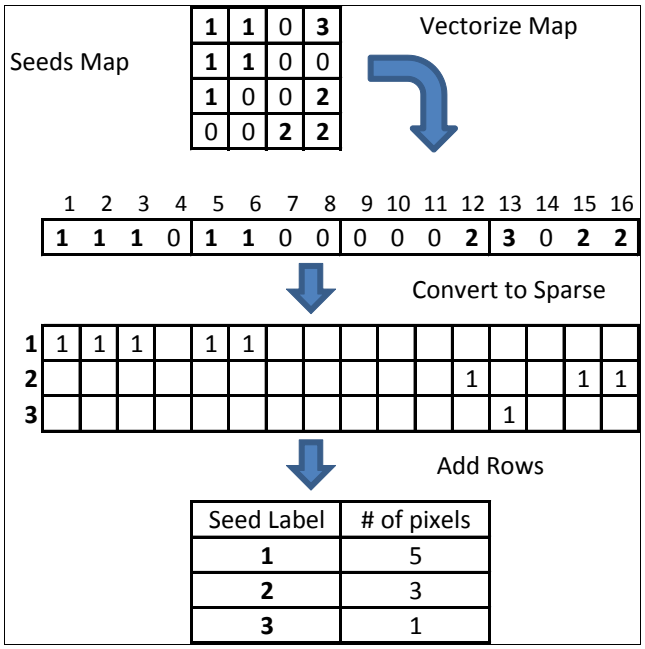

Fig. 3.7: Method to identify number of pixels per seed.

Parent-child similarity is evaluated by computing the distance between the mean L^*, a^* and b^* values of *PSs* and *CSs*, $(m_{L^*_{PSs}}, m_{a^*_{PSs}}, m_{b^*_{PSs}})$ and $(m_{L^*_{CSs}}, m_{a^*_{CSs}}, m_{b^*_{CSs}})$ respectively, in the Euclidean space (E^3), defined as:

$$D_E^{PSs-CSs} = \sqrt{(m_{L^*_{PSs}} - m_{L^*_{CSs}})^2 + (m_{a^*_{PSs}} - m_{a^*_{CSs}})^2 + (m_{b^*_{PSs}} - m_{b^*_{CSs}})^2} \quad (3.12)$$

The combination of the CIE $L^*\,a^*\,b^*$ color space and the Euclidean distance metric was employed because: 1) it assures the comparison of colors is similar to the differentiation of colors by the human eye, 2) the increased complexity of a different distance metric like the Mahalanobis distance does not improve the results, due to the small variance of the regions being compared, owed to their spatial proximity. The maximum color distance to allow the integration of a child seed to its parent was empirically chosen as 10. Thus, the adjacent CSs whose mean $L^*\,a^*\,b^*$ values are found comparable to corresponding parents are merged with them while the rest are discarded, as portrayed in Fig. 3.5. The aforementioned procedure continues until the gradient threshold T_n is reached (see Fig. 3.6(h)), which signals the addition of new seeds to the existing seed map, in a procedure termed as dynamic seed addition.

The dynamic seed addition procedure is responsible for the addition of new seeds in unsegmented image areas that are non-adjacent to the existing seeds in the same interval (T_{n-1} to T_n) that was previously utilized for region growth. All pixels in the gradient map whose value belong to the interval from T_{n-1} to T_n, that are nonadjacent to existent seeds and larger than the *MSS* criterion (as shown in Fig. 3.6(i)) are added to the current seed map. The resultant ensemble of seeds, become *PSs* map for the succeeding *RGI* (see Fig. 3.6(j)). In general, it was found that seeds whose L^* a^* b^* values greatly differ from all existing regions remain as independent entities throughout the region growth and dynamic seed addition phase.

In case of natural scene images where gradient content can dramatically vary, accomplishing region growth in the aforementioned iterative procedure over the entire gamut of gradient values present in these images can be computationally intensive. To this effect, the totality of the growth procedure in the proposed algorithm was restricted to a finite number of *RGIs* which may span only a portion of the total gamut of gradient values in an image, but sufficient enough to segment a large portion of it. This limit on the number of *RGIs* was chosen based on the average percentage of total segmented image area, determined utilizing 300 natural scene images provided by the University of California at Berkeley. We found that with a growth factor ($n\Delta g$ defined in Section 2.3) varying from 10% to 50% obtained utilizing $1 \le n \le 5$ and $\Delta g = 10\%$, on an average more than 85% area of an image to be segmented, and hence constrain the growth phase to a maximum of five *RGIs* (*N=5*), as shown in Fig. 3.8(a). This constraint on the number of *RGIs*, results in a small portion of the image largely comprising of regions of color transitions in the periphery of existing seeds being left unsegmented, at the conclusion of the growth procedure, as shown in Fig. 3.8(b). These unsegmented regions are assigned labels in a procedure known as residual region growth that involves local neighborhood-based mode filtering and morphological dilation operations. Mode filtering is a technique in which un-labeled pixel locations *(i,j)* surrounded by existing seeds in their respective local 3X3 neighborhood (β), are assigned the most frequently occurring label from amongst the non-zero elements of that neighborhood (β_{nz}), using a non-linear spatial mode filter (m_f) defined as:

$$m_f(i,j) = \begin{cases} mode(\beta_{nz}) & if \quad mode(\beta_{nz}) \ is \ unimodal \\ \phi : \phi \in mode(\beta_{nz}) & if \quad mode(\beta_{nz}) \ is \ multimodal \end{cases} \tag{3.13}$$

where, $$\beta_{nz} = \{\beta(m_1, m_2) : \beta(m_1, m_2) > 0, \ m_1 \in [i-1, i+1], m_2 \in [j-1, j+1]\} \tag{3.14}$$

In locations where the mode of β_{nz} is not unique (multimodal), a random label assignment ϕ from the acquired multiple mode values is performed, as represented in Eq. (3.13). At this stage the pixels that remain unsegmented are the ones whose corresponding local neighborhoods do not constitute any of the existing seeds (see Fig. 3.8(c)). To this effect, an iterative morphological label assignment is employed,

where-in all existing seeds are repeatedly dilated using a 3X3 structuring element ψ (defined previously) until there exists no unassigned pixels, to yield the final region growth map, as displayed in Fig. 3.8(d). It is important to understand that as an alternative to residual region growth, the iterative growth process (previously described) can be continued until all pixels in the image have assigned labels, to generate the final region growth map. The aforementioned choice of using the residual region growth procedure does not intervene with the intention of the work presented, since the goal was to ensure that every pixel in input image had an assigned label without any significant increase in computational overhead for the last few unsegmented pixels in the image. The RGDSA phase has been largely implemented in a vectorized approach to achieve high levels computational efficiency.

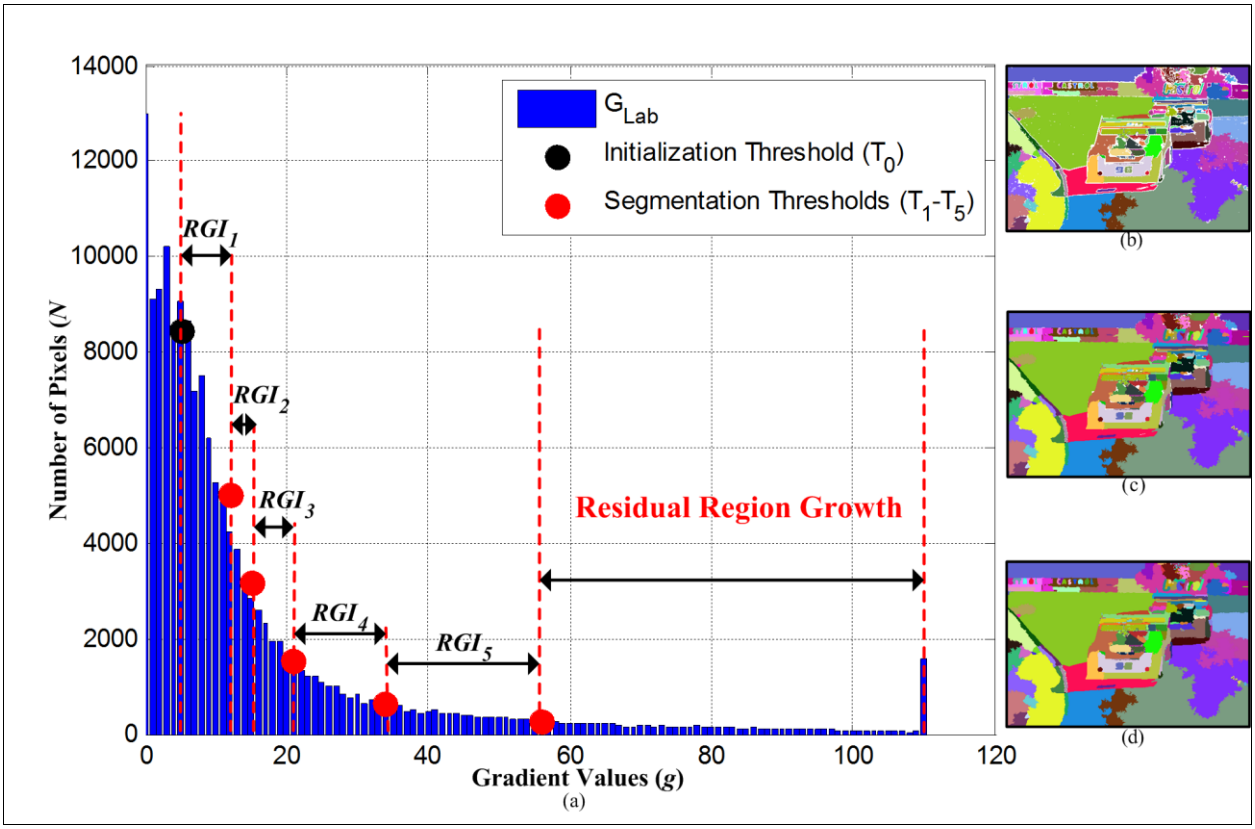

Fig. 3.8: (a) Gradient histogram of the 'Cars' image with adaptively generated segmentation thresholds for $1 \leq n \leq 5$ and $\Delta g = 10\%$, (b) seed map at the end of five *RGIs*, (c) neighborhood label assignment, (d) iterative morphological label assignment to yield final region growth map.

3.6. Texture Characterization

In case of natural scene imagery, the segmentation task is often hampered by the presence of regions/patterns composed of multiple shades of colors or intensity variations due to surface/material properties like density, gradient, coarseness, directionality etc, to name a few. Such regions/ patterns are referred to as 'textures' and are broadly classified into structured and stochastic types, in the image understanding domain. Structured textures are often man-made and have regularity in their appearance, such as a brick wall, interwoven fiber etc, while stochastic textures are natural and are completely random patterns, such as leopard skin, tree bark, grass etc. Due to the extensive presence of such patterns in natural scene images the GSEG algorithm has been equipped with a texture characterization module (M3 in Fig. 3.1), which characterizes different textures, in terms of the average information provided by intensity variations present in distinct image regions.

A fundamental principle in information theory is based on the hypothesis that the presence of information can be modeled as a probabilistic process, and that the amount of information contained in a random event is inversely proportional to the probability of the occurrence of that event [Cover and Thomas, 1991]. Thus, if $\{x_1, x_2, \ldots, x_J\}$ are a set of random gray levels present in an image, and $\{P(x_1), P(x_2), \ldots, P(x_J)\}$ are the corresponding probabilities of occurrences of each of these gray levels, an arbitrary gray level $\{x_i\}$ from the set is said to contain:

$$I(x_i) = \log_2 \frac{1}{P(x_i)} = -\log_2 P(x_i) \tag{3.15}$$

binary units or bits of information when the base of the logarithm is 2. Furthermore, for an image comprising of k pixels, the law of large numbers states that a gray level $\{x_i\}$ exists on average of $kP(x_i)$ times. Consequently, the total information content (I) in these k pixels, whose intensity values is modeled as a discrete random variable X, is given by:

$$I(X) = -kP(x_1)\log_2 P(x_1) - kP(x_2)\log_2 P(x_2) \ldots - kP(x_J)\log_2 P(x_J) = -k\sum_{i=1}^{J} P(x_i)\log_2 P(x_i) \tag{3.16}$$

Therefore, the average information content per pixel is given by:

$$H(X) = -\sum_{i=1}^{J} P(x_i)\log_2 P(x_i) \tag{3.17}$$

Fig. 3.9: Texture characterization: (a) I_{RGB}, (b) I_{Lab}, (c) indexed image, (d) texture channel.

Apart from information content, the quantity $H(X)$ also symbolizes the degree of randomness present in the image, and is popularly known as entropy. The entropy calculation in Eq. (3.17) defined for a single random variable (single channel gray image) can be extended to multiple random variables *X, Y, Z* (three channel color image) by computing the joint entropy, defined as:

$$H(X,Y,Z) = -\sum_{i}\sum_{j}\sum_{k} P(x_i, y_j, z_k)\log_2 P(x_i, y_j, z_k) \tag{3.18}$$

However, in order to generate an entropy-based texture descriptor with minimal computational requirements, we exploit the information presented in the L^*, a^* and b^* channels without computing the joint entropy, by first uniformly quantizing these channels into 6^3 or 216 distinct levels (shown in Fig. 3.9(c)), through division of the 8-bit encoded $L^*\ a^*\ b^*$ data cube into small boxes, followed by a mapping all information that falls within each box to the mean value at the center of that box. To this effect, each pixel of an image is subsequently indexed to one of these 216 levels, essentially reducing the probability of the occurrence of each level to a one-dimensional random variable. Furthermore, the advantage of quantizing the $L^*\ a^*\ b^*$ cube over the RGB color cube is that, unlike uniform $L^*\ a^*\ b^*$ data, if non-uniform RGB data is uniformly quantized, a constant distance between and any two quantization levels will result in large variation of perceptual color difference [Chou and Wu, 2003]. Finally a texture channel (see Fig. 3.9(d)) is created by local neighborhood-based entropy calculations wherein, the entropy in a 9-by-9 local neighborhood around every pixel of the indexed image is computed and the resultant value is assigned to the central pixel location of the corresponding neighborhood.

3.7. Region Merging using a Multivariate Analysis

The region growth and dynamic seed addition procedure illustrated in the previous sub-section, in general, was performed primarily based on the similarity of $L^*\ a^*\ b^*$ data between image regions. Consequently, the region growth map obtained at the end of this procedure, in general comprises of over-segmented image regions due to illumination variations, occlusions, texture disparities etc. We incorporate an effective method to analyze grouped data from the statistical field, to merge all over-segmented regions. This method better known as a multivariate analysis allows us to take regions that

have been separated due to occlusion, or small texture differences, and merge them together. Thus, we employ a region merging module (M4 in Fig. 3.1) that fuses color and texture information, to merge over-segmented regions as deemed necessary, yielding the final segmentation output. However in order to facilitate the aforementioned task, a multivariate analysis of all independent regions utilizing their corresponding L^*, a^*, b^* and texture information, is carried out based on the procedure [Krzanowski., 1988] described in the following sub-section. The essence of this method is to investigate the possibility that multiple groups/regions with various features can be associated with a single factor that enables them to be merged together.

3.7.1 Multivariate ANalysis Of Variance (MANOVA)

The Multivariate Analysis Of Variance abbreviated as MANOVA, is a popular statistical method employed in highlighting differences between groups of data [Krzanowski, 1988], cumulatively structured in the form of a matrix of dimensions $n \times p$, in which n samples are divided into g groups, where each sample is associated with p variables $x_1, x_2, \ldots, x_p$. To this effect, the goal of the MANOVA procedure is to find the optimal single direction in the p-dimensional space, so as to conveniently view differences between various groups.

In the general case of p variables, any direction in the p-dimensional space can be designated as a linear combination of certain vectors $(a_1, a_2, \ldots, a_p)$, which can be utilized to convert every p-variate sample to a univariate observation $y_i = a^T x_i$, where $a^T = (a_1, a_2, \ldots, a_p)$. However since the n data samples are divided into g groups, the obtained univariate observations are re-labeled as y_{ij} denoting the y value for the j^{th} sample in the i^{th} group, where $i = 1 \ldots g$ and $j = 1 \ldots n_j$. In order to establish whether the aforementioned univariate observations demonstrate differences between groups, the total sum-of-squares of y_{ij} is partitioned into its Sum-of-Squares Between (SSB)-groups and Sum-of-Squares Within (SSW)-groups components, defined as:

$$SSB(a) = \sum_{i=1}^{g} n_i (\overline{y_i} - \overline{y})^2 \qquad \text{and} \qquad SSW(a) = \sum_{i=1}^{g} \sum_{j=1}^{n_j} (y_{ij} - \overline{y_j})^2 \tag{3.19}$$

Here, $\overline{y_i} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{ij}$, $\overline{y} = \frac{1}{n} \sum_{i=1}^{g} \sum_{j=1}^{n_i} y_{ij} = \frac{1}{n} \sum_{i=1}^{g} n_i \overline{y_i}$ and the notation (a) in Eq. (3.19) is utilized to underscore the fact that the SSB and SSW components vary with the choice of a. Utilizing these components a mean square ratio (F), to highlight group differences, is obtained as:

$$F = \left\{ \frac{1}{(g-1)} SSB(a) \right\} \Big/ \left\{ \frac{1}{(n-g)} SSW(a) \right\} \tag{3.20}$$

From Eq. (3.20) it can be seen that the larger the value of F, the more variability exists between groups than within groups. Consequently, the optimal choice of the coefficients $a^T = (a_1, a_2, \ldots, a_p)$ will be the one that yields the largest value for F. However, to ascertain the optimal values of a, multivariate analogues of the between-groups and within-groups sum of squares components used in the univariate analysis of the variance are computed, and defined as:

$$B_0 = \sum_{i=1}^{g} n_i (\overline{x_i} - \overline{x})(\overline{x_i} - \overline{x})^T \quad \text{and} \quad W_0 = \sum_{i=1}^{g} \sum_{j=1}^{n_j} (x_{ij} - \overline{x_j})(x_{ij} - \overline{x_j})^T \tag{3.21}$$

where B_0 known as the between-groups sum-of-squares and products matrix and W_0 known as the within-groups sum-of-squares and products matrix, should be positive definite matrices. Furthermore, the notations x_{ij}, $\overline{x_i}$, $\overline{x}$ are analogous to y_{ij}, $\overline{y_i}$, $\overline{y}$ respectively. Since $y_{ij} = a^T x_{ij}$ Eqs. (3.19), and (3.20) can be re-written as:

$$SSB(a) = a^T B_0 a \quad \text{and} \quad SSW(a) = a^T W_0 a \tag{3.22}$$

$$F = \left\{ \frac{1}{(g-1)} SSB(a) \right\} \Big/ \left\{ \frac{1}{(n-g)} SSW(a) \right\} = \left\{ \frac{1}{(g-1)} a^T B_0 a \right\} \Big/ \left\{ \frac{1}{(n-g)} a^T W_0 a \right\} = \frac{a^T B a}{a^T W a} \tag{3.23}$$

where $B = B_0/(g-1)$ and $W = W_0/(n-g)$ are the between-groups and within-groups covariance matrices respectively.

The choice of the coefficients $a^T = (a_1, a_2, \ldots, a_p)$ which maximizes the value of F in Eq. (3.23) signifies the optimal single direction (or the best linear combination $y = a^T x$) in the p-dimensional space, so as to highlight differences between various groups, and can be obtained by differentiating Eq. (3.23) with respect to a and assigning result to zero. To this effect, we have:

$$Ba - \left(\frac{a^T B a}{a^T W a} \right) W a = 0 \;\Rightarrow\; Ba - lWa = 0 \;\Rightarrow\; (B - lW)a = 0 \;\Rightarrow\; (W^{-1}B - lI)a = 0 \tag{3.24}$$

where $a^T Ba / a^T Wa = l$ is a constant, equal to the maximum value of the mean square ratio (F). Also, for Eq. (3.24) to be satisfied, it can be inferred that l must be an Eigenvalue and a must be an Eigenvector of $W^{-1}B$. Moreover, since l is constant value at the maximum of F, a must be Eigenvector associated with the largest Eigenvalue of $W^{-1}B$, that determines the optimal linear combination $y = a^T x$. Note that for a distinct separation of groups (greater variability between groups than within groups), l will be significantly greater than unity.

When the number of groups (g) or the dimensionality of the original space (p) is large, the goal of determining a single direction in the p-dimension space renders an inefficient solution to view disparities between various groups. However, since Eq. (3.24) often possesses more than one solution, multiple differentiating directions can be generated, whose efficiency in delineating groups of data depends on the magnitude of the Eigenvalue/Eigenvector pairs. To this effect, if Eq. (3.24) possessed s non-zero Eigenvalues $(l_1, l_2, \ldots, l_s)$ with corresponding Eigenvectors $(a_1, a_2, \ldots, a_s)$, a set of new variates $(y_1, y_2, \ldots, y_s)$ known as canonical variates can be obtained according to $y_i = a_i^T x_i$, and the space spanning all y_i's is termed as a canonical variate space. Following this, Eq. (3.24) can be re-written in matrix terms as $BA = WAL$, where A is matrix of all a_i's of dimensions $(p \times s)$, while L is matrix of all l_i's of dimensions $(s \times s)$. Furthermore, in this space, the mean of an arbitrary i^{th} group of individuals can be represented as $\overline{y_i} = A^T \overline{x_i}$.

An appropriate measure to quantify the variability between two random groups (i^{th} and j^{th}) of data is the distance between the corresponding group means, given by:

$$D = ((\overline{x_i} - \overline{x_j})^T M (\overline{x_i} - \overline{x_j}))^{1/2} \Rightarrow D^2 = ((\overline{x_i} - \overline{x_j})^T M (\overline{x_i} - \overline{x_j})) \tag{3.25}$$

Here M is a matrix that modifies the influence of each variate in the aforementioned distance computation. Moreover, to exploit the covariance's between variables and as well as differential variances, M can be chosen to be the inverse of the Within-groups dispersion matrix (W). The resultant distance measure for this choice of M yields the Mahalanobis distance, defined as:

$$D = ((\overline{x_i} - \overline{x_j})^T W^{-1} (\overline{x_i} - \overline{x_j}))^{1/2} \Rightarrow D^2 = ((\overline{x_i} - \overline{x_j})^T W^{-1} (\overline{x_i} - \overline{x_j})) \tag{3.26}$$

The Euclidean between i^{th} and the j^{th} groups in the canonical variate space after substitution for $\overline{y_i}$ and $\overline{y_j}$, can be written as:

$$D = ((\overline{y_i} - \overline{y_j})^T (\overline{y_i} - \overline{y_j}))^{1/2} \Rightarrow D^2 = ((\overline{x_i} - \overline{x_j})^T AA^T (\overline{x_i} - \overline{x_j})) \tag{3.27}$$

Furthermore, it can be shown that $AA^T = W^{-1}$ [Krzanowski, 1988], resulting in Eq. (3.27) being equal to Eq. (3.26). Thus, by generating a canonical variate space in a manner described in this section, the Euclidean distance between the group means in this space is equivalent to the Mahalanobis distance in the original p-dimension space. Moreover, since the Mahalanobis distance metric takes into consideration the covariance and differential variance between variables, this distance measure is utilized to measure of variability between two multivariate populations.

3.7.2 Region Merging using MANOVA

As mentioned previously the region merging module is integrated with the MANOVA procedure (see Fig. 3.10), to analyze data associated with each group in the region growth map (generated previously), to produce output segmentations that are spatially and spectrally coherent with the content of image being segmented. Consequently, to facilitate the aforementioned MANOVA-based region merging methodology, at the commencement of processing in this module, the L^*, a^*, b^* and texture data associated with each group in the region growth map are vectorized and concatenated to matrix of dimensions equivalent to the total number of pixels in the image and number of variables (L^*, a^*, b^*, texture) per pixel. The result matrix is employed in the MANOVA procedure involving the Mahalanobis distance (or similarity value) calculation between all possible group pairs, to identify and merge groups with similar characteristics.

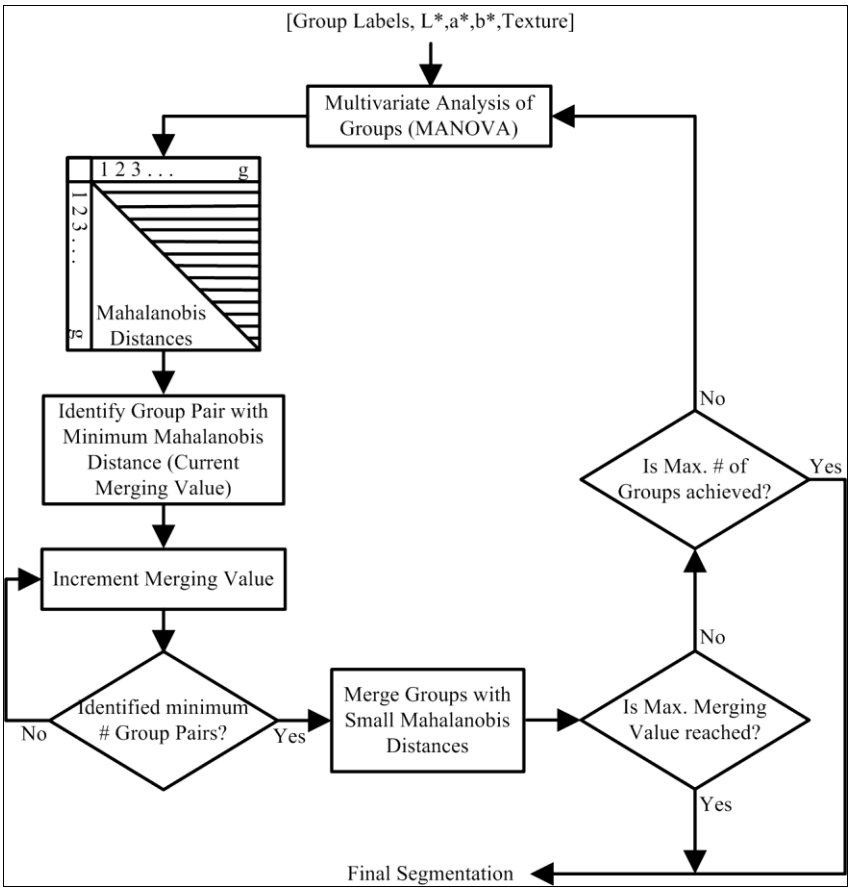

Fig. 3.10: Region merging module (M4) using MANOVA.

The merging process is commenced by identifying the pair of groups with the minimum Mahalanobis distance, signifying the maximum similarity. However in order to reduce the number of iterations of the merging protocol for computational efficiency, by avoiding the merging of only a single group pair per

iteration, the obtained distance value between the two most similar groups is gradually increased until a larger set of similar groups pairs (empirically set at five) are obtained, as depicted in Fig. 3.10. Subsequently, the acquired group pairs are merged with each other from the most similar group pair of the set, to the least similar one, eventually concluding a single iteration of the merging process. Following this, the Mahalanobis distances is recomputed for all possible group pairs comprised in the new segmentation map, and the process is repeated until either a desired number of groups (S) is achieved or the smallest distance between groups is larger than a certain threshold (T) between two arbitrary groups, as portrayed in Fig. 3.10. These termination criterions ensure that that all images displayed a similar level of segmentation, and were empirically chosen to be 50 and 2 respectively. However these could be varied depending on the application for which this algorithm is being used.

3.8 Summary

From the discussion in this chapter it can be seen that the novelty of the proposed segmentation methodology relative to prior art to overcome some of the existent shortcomings is twofold. Firstly, our region growing/formation process: (a) simultaneously lays emphasis on the homogeneous and heterogeneous characteristics of a scene (via spectral and gradient information), (b) does not requisite manual selection of seeds, (c) refrains from pixel-by-pixel agglomeration, (d) avoids processing order dependency of regions, and (e) inherits a parallel design (grows multiple image regions simultaneously). Secondly, our region merging approach facilitates refinement of regions based on analysis of a multi-attribute space to fuse information from distinct features to define meaningful spatial segments, even in the presence of texture disparities, spatial separation and occlusion. The following chapter discusses a multi-resolution extension of the proposed algorithm with the objective of handling sub-sampled versions of image and/or producing multiple segmentations representing distinct levels of detail with robustness to scalability as well as to explore the possibility of facilitating a potential solution that computationally measures up to meet the demands of commercial applications involving segmentation.

Chapter 4

Multiresolution Adaptive and Progressive Gradient based Color Image Segmentation

The proposed Multiresolution Adaptive and Progressive Gradient-based color image SEGmentation (MAPGSEG) algorithm, embodied in six modules is shown in Fig. 4.1. The first module (M1) performs dyadic wavelet decomposition for multi-resolution representation of the input image. The second module (M2) is utilized to perform the functionalities of color space conversion, gradient computation and adaptively generate thresholds employed in distinct stages of region processing at varied resolution levels of the input image pyramid. The third module (M3) carries out a progressively thresholded region growth procedure involving distributed dynamic seed addition. The fourth module (M4) is responsible for identifying transferable regions from one resolution to another by exploiting the interim results as a-priori information. Texture modeling is implemented in the fifth module (M5). The proposed algorithm culminates in a region merging module (M6) to yield interim segmentations at low resolutions, and the final segmentation output at the highest resolution. The following sub- sections elucidate each of these modules in detail.

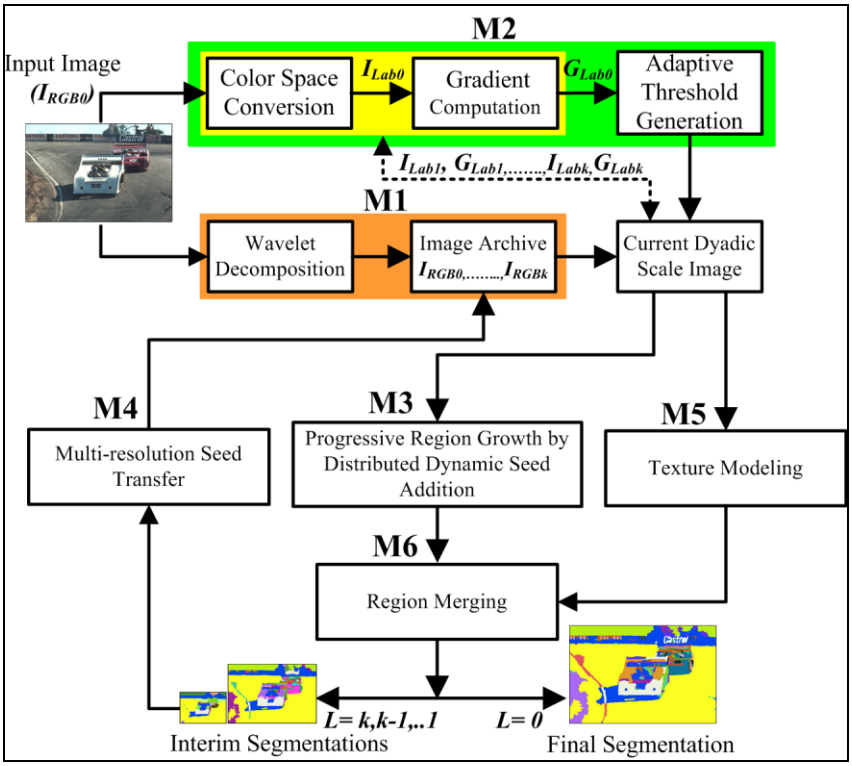

Fig. 4.1: Overview of the MAPGSEG algorithm.

4.1. Dyadic Wavelet Decomposition

The first module (M1) performs a dyadic (powers of 2) wavelet decomposition scheme [Mallat., 1989], utilizing the Daubechies 9/7 analysis coefficients employed in the JPEG2000 compression methodology [Christopoulos et al., 2000], to obtain high-quality approximations of the input RGB image (I_{RGB0} of dimensions d_{M0} by d_{N0}) at different resolution levels. These levels are designated as $L=0, 1, 2,\ldots, k$ for a k-level decomposition, as described in Fig. 4.2(a).

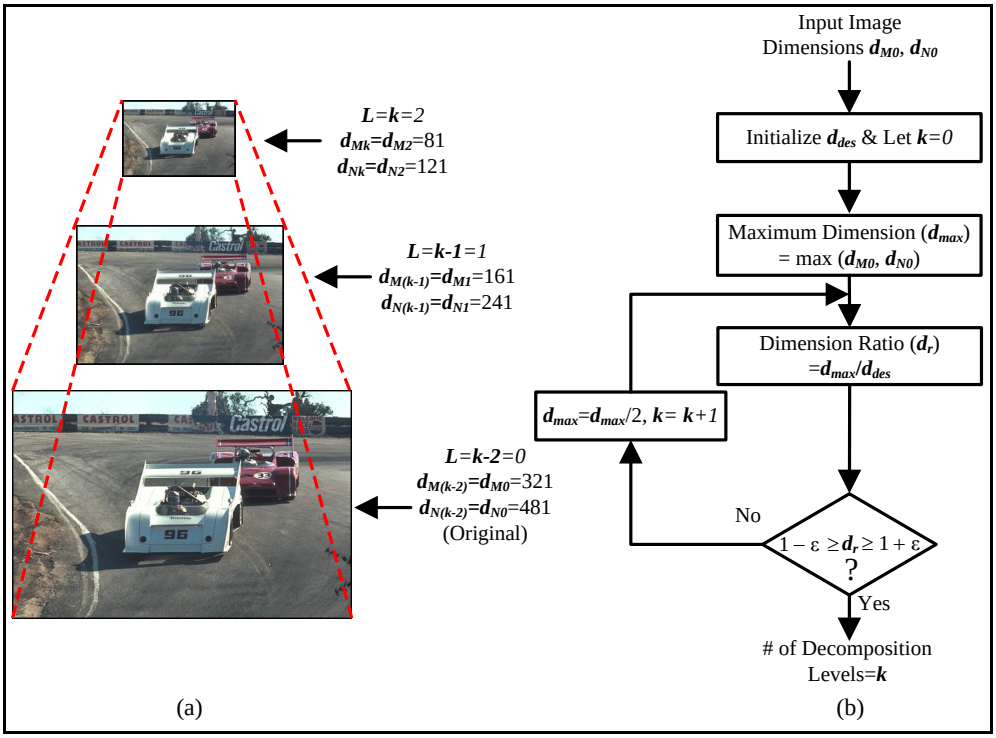

Fig. 4.2: (a) Two level decomposition (k=2) with corresponding designations, (b) Automatic determination of number of decomposition levels based on smallest desired dimension (d_{des}).

The number of decomposition levels 'k' is dynamically determined for a randomly selected image based on a user or application defined parameter called 'desired dimension' (d_{des}) , using the procedure portrayed in Fig. 4.2(b). This parameter is defined as the smallest workable dimension desired by a user or constrained by an application. Since often applications can be restricted by the smallest size of an image that they can handle, the MAPGSEG is designed with the flexibility to allow the application or user the option to set the smallest workable dimension for segmentation. Once d_{des} is initialized (arbitrarily set at 128), and the maximum dimension (d_{max}) of I_{RGB0} is found, the dimension ratio (d_r) of d_{max} to d_{des} is computed. In the general case where d_{des} is chosen such that $d_{des} < \{d_{M0}, d_{N0}\}$, the dimension ratio will be greater than 1. To this effect, the value of d_{max} is reduced by a factor of 2 and d_r is recomputed. The procedure is repeated until d_r is in the vicinity of 1 (that is $1 \pm \varepsilon$ where $\varepsilon << 1$),

consequently the number of iterations (k) indicates the number of dyadic decomposition levels that will result in the d_{max} being in the vicinity of d_{des} (since d_{des} may or may not be a dyadic scale of d_{max}). Furthermore, the smaller of the two image dimensions is automatically mapped such that aspect ratio of the image at every scale is the same. Having obtained the value of k as illustrated above, the color components of I_{RGB0} are independently decomposed to the smallest resolution (I_{RGBk}), simultaneously fusing corresponding channel information at all k levels, to form an 'image archive' (I_{RGB0}, I_{RGB1},.., I_{RGBk}), as depicted in Fig. 4.1. Hence, the decomposition protocol is performed only once in our work, without having to be repeated prior to execution at every resolution level.

4.2. Color Space Conversion and Gradient Computation

The second module (M2), is initiated by a color conversion of the original input (I_{RGB0}) from RGB to CIE $L^* a^* b^*$ (see Fig. 4.1) for reasons specified in Section 3.1. Following this, the magnitude of the gradient (G_{Lab0}) of I_{Lab0} is computed utilizing the algorithm described in Section 3.2. The resultant gradient information is employed in an adaptive threshold generation scheme to obtain a set of segmentation thresholds.

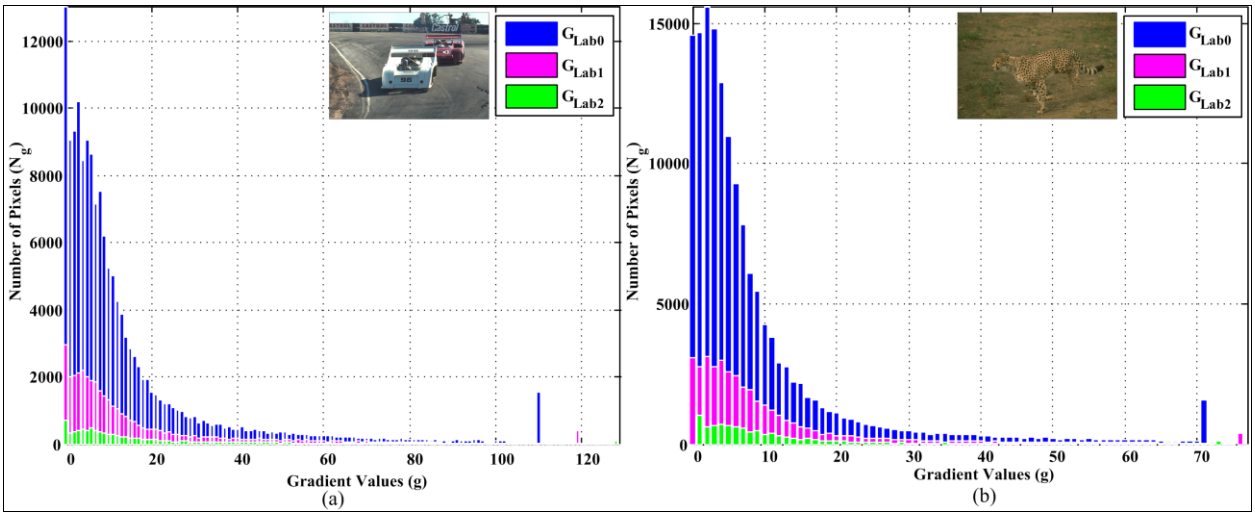

Fig. 4.3: Comparison of gradient information using CIE $L^* a^* b^*$ data at different resolution levels of: (a) Cars, (b) Cheetah.

Similar to the GSEG algorithm, the MAPGSEG framework is designed such that it is iterative and dynamic over the gamut of gradient values of the input image. Consequently, apart from generating a set of suitable segmentation thresholds, the aforementioned color conversion scheme was also incorporated to facilitate a reduction in computational requirements, as gradient values computed in the CIE $L^* a^* b^*$ were found to be confined to a significantly smaller range than their RGB equivalents, as

discussed in Section 3.1. Furthermore, since gradient content in an image gradually increases from coarse to fine resolutions with nearly similar density distributions (as depicted in Figs. 4.3(a) and 4.3(b)), the aforementioned segmentation thresholds computed utilizing G_{Lab0} were employed unaltered for processing at all resolution levels.

4.3. Adaptive Threshold Generation

In contrast to the GSEG algorithm, region processing in the MAPGSEG algorithm is carried out in three distinct phases, namely: 1) Initial Clustering, that instigates region formation by identifying potential homogeneous sites where seeds are placed, 2) Progressive Region Growth by Distributed Dynamic Seed Addition, required for the growth of existent seeds to their maximum extent, accompanied by the addition of new seeds in regions that remain unsegmented at specific stages of processing, and 3) Multi-resolution Seed Transfer, responsible for identifying seeds transferable between resolutions. The three phases are carried out by thresholding the histogram of the gradient map, commencing in low gradient (or color homogeneous) regions and culminating in strong gradient (color transition) regions.

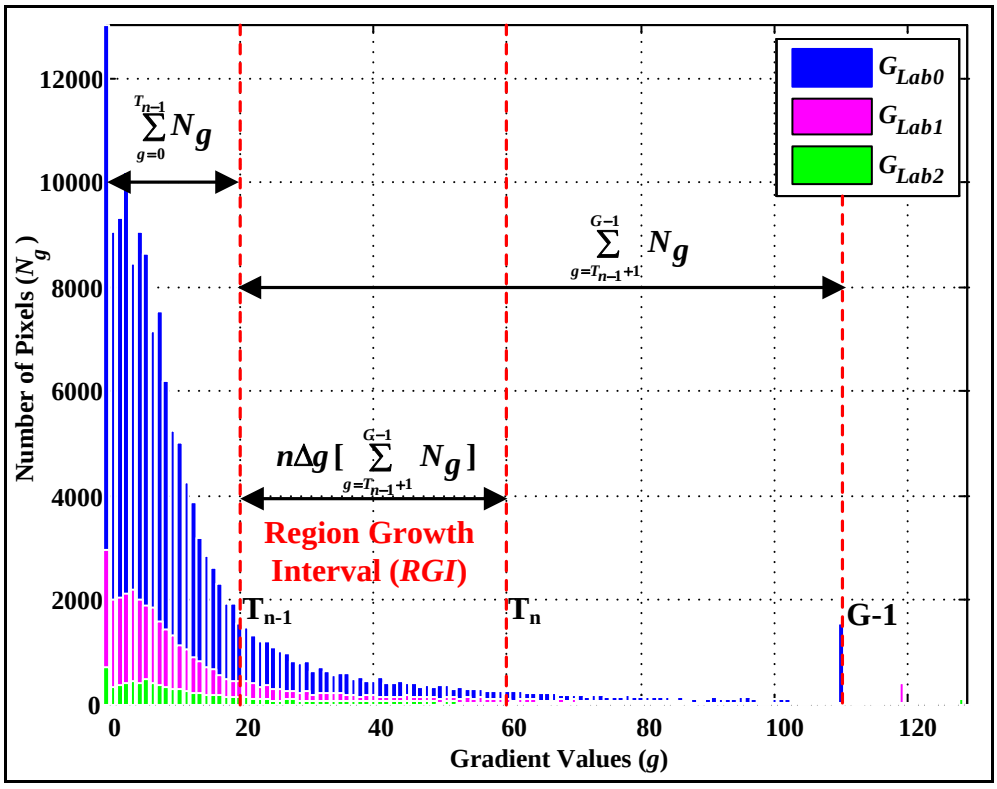

Fig. 4.4: Adaptive threshold generation.

The primary objective of the adaptive threshold generation scheme is to provide suitable thresholds to delineate the instants at which seed addition should be performed during region growth as well as demarcate various gradient quantization levels that are utilized in the seed transfer procedure, at different

resolution levels (in a procedure described later). Furthermore, since gradient content gradually increases from low to high resolutions with nearly similar density distributions (as mentioned in Section 4.2), the aforementioned segmentation thresholds T_1 to T_n differentiating various Region Growth Intervals as well as gradient quantization levels were adaptively computed utilizing G_{Lab0} in a manner described in Section 3.3 (see Fig. 4.4 which depicts the threshold generation scheme for multiresolution data, analogues to Fig. 3.3 for single scale data).

4.4. Initial Clustering

Region formation is initiated (see Fig. 4.5) by identifying potential homogeneous seed sites utilizing an initialization threshold (T_0) in extremely low gradient ranges (e.g. T_0=5). All the pixels in the gradient map that fall below T_0 undergo connected component labeling, to generate a map comprising of initial seeds or Parent Seeds (*PSs*) as shown in Fig. 4.5(d). The generation of *PSs* concludes the initial clustering phase which is performed only at the lowest resolution level (L=k).

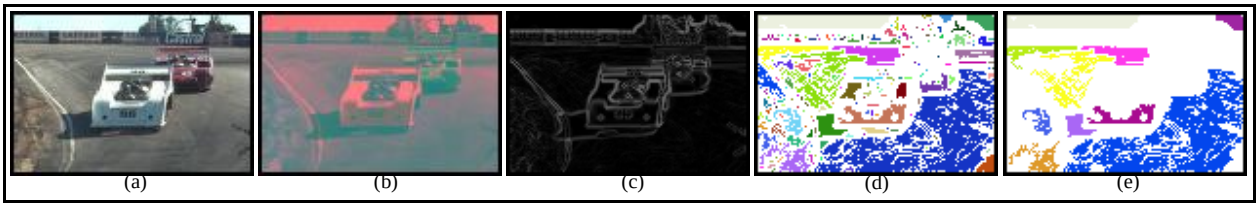

Fig. 4.5: Initial Clustering phase for a 2-level decomposition: (a) Lowest resolution image I_{RGB2}, (b) I_{Lab2}, (c) G_{Lab2}, (d) *PSs* map, (e) *PSs* map after employing spatial constraints.

However, to prevent the generation of multiple clusters within homogeneous and connected regions, *PSs* are constrained to clusters of pixels which are of certain minimum spatial extent (see Fig. 4.5(e)). Spatial constraints (of seed extent) required during various stages of region formation are chosen proportional to a Minimum Seed Size (*MSS*) criterion that defines the minimum desired size of an independent seed. The *MSS* criterion for region processing at an arbitrary L^{th} decomposition level (MSS_L) is chosen such that: 1) it is very small in comparison to the corresponding L^{th} level image area (that is $MSS_L << \{d_{Mk} d_{Nk}\}$) to ensure that fine details are captured in the segmentation process, as well as 2) it is a function of the down sampling rate 2^L employed during decomposition to warrant the processing of seeds of 'meaningful' sizes in comparison to corresponding L^{th} scale image area. Consequently, based on the aforementioned requirements the MSS_L criterion is computed as:

$$MSS_L = 2^L \alpha \{d_{ML} d_{NL}\} \tag{4.1}$$

where α is a small percentage (typically 0.01%) of L^{th} scale image area. Hence, for the initial clustering phase performed only at the lowest resolution L=k^{th} level (previously mentioned), a Minimum Seed Size criterion proportional to $MSS_k = 2^k \alpha \{d_{Mk} d_{Nk}\}$ is employed.

4.5. Progressive Region Growth by Distributed Dynamic Seed Addition

A flow chart of the region growth module (M3) within the MAPGSEG framework is shown in Fig. 4.6.

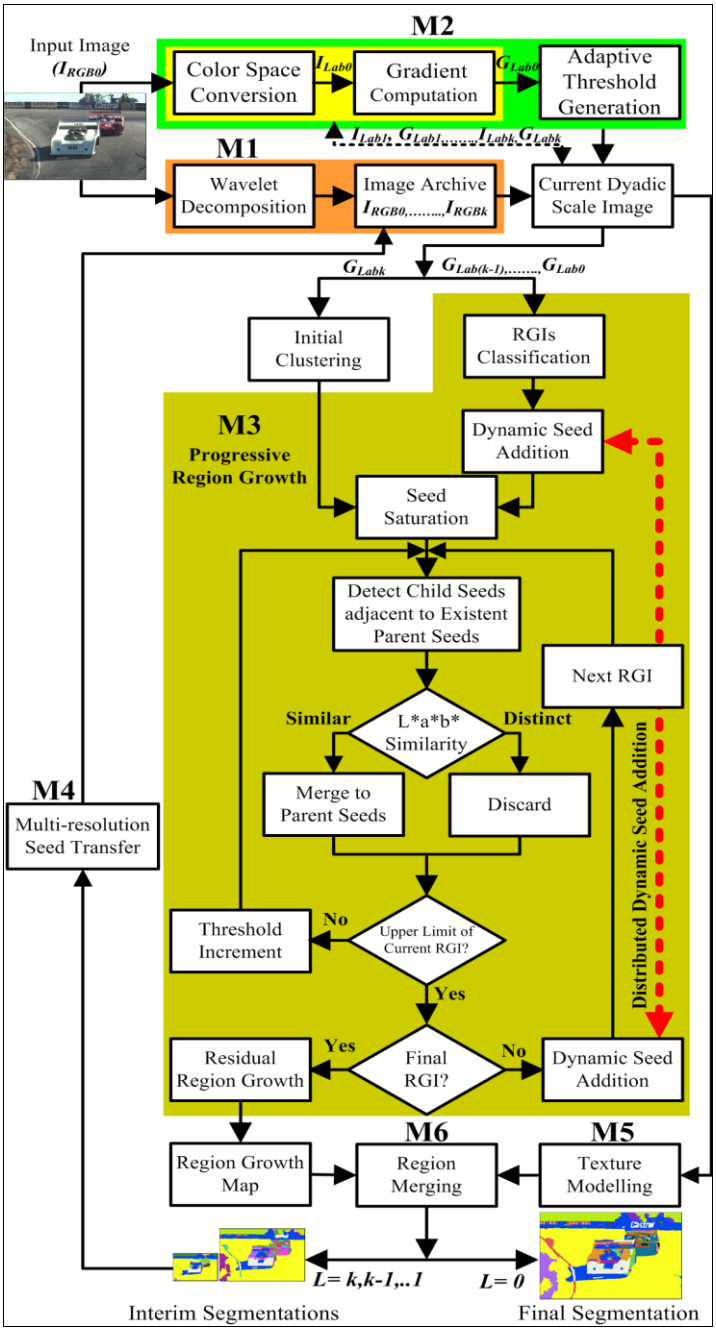

Fig. 4.6: Progressive Region Growth utilizing Distributed Dynamic Seed Addition

The initial clustering stage that instigates region formation yields a Parent Seeds (*PSs*) map (Fig. 4.5(e) or 4.7(a)). To ensure an efficient growth procedure that channelizes computational costs to segment meaningful regions, prior to region growth all *PSs* are subjected to a seed saturation process, where-in small/isolated unsegmented regions enclosed within *PS* boundaries are directly assigned the labels of their

corresponding parents. Consequently, the resultant 'saturated' seed map (see Fig. 4.7(d)) facilitates the growth of seeds to occur in the outward direction or towards large unsegmented image areas, rather than in small unsegmented seed interiors that will visually not have any impact in the final segmentation result. Subsequently, an iterative region growth procedure (described in Section 3.5), is performed over all *RGIs* (summarized in this section for completeness).

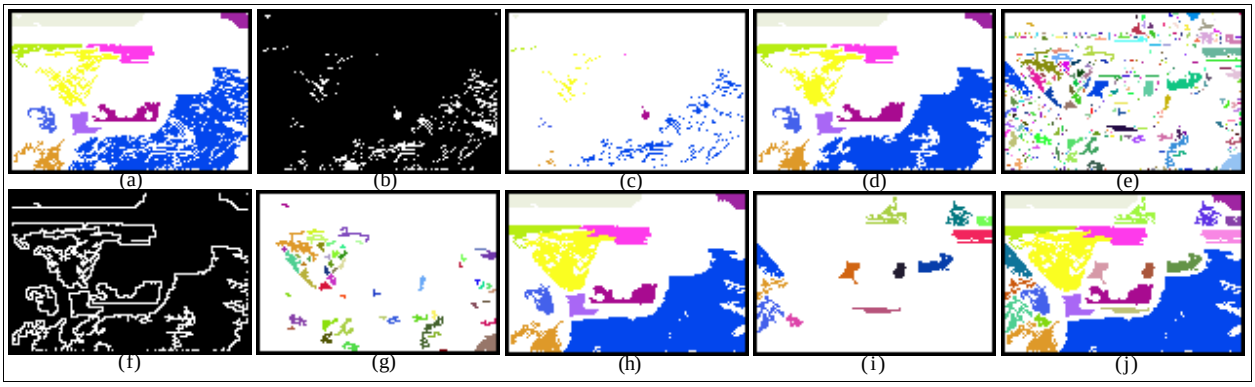

Fig. 4.7: (a) *PSs* map, (b) small/isolated unsegmented regions enclosed within *PS*s, (c) direct label assignment of these isolated regions, (d) *PSs* map after seed saturation, (e) C*Ss* map, (f) boundaries of saturated *PSs* map, (g) Adjacent *CSs*, (h) seed map at the end of 1st RGI, (i) new seeds generated during dynamic seed addition, (j) *PS* map for 2nd RGI.

In an arbitrary (n^{th}) RGI spanning from T_{n-1} to T_n, the growth threshold (which is a gradient value) is sequentially incremented, detecting new areas which are referred to as Child Seeds (*CSs*). However, only the *CSs* that are adjacent to existent *PSs* are administered at this point. Adjacent *CSs* whose mean L*a*b* values are found comparable to corresponding parents, are directly merged with them and the rest are discarded. This process continues until the gradient threshold T_n is reached, which signals the addition of new seeds to the existing seed map. The resultant seed map yields the *PSs* map for the $(n+1)^{th}$ *RGI*. Regions that are left unsegmented at the end all *RGIs* are morphologically assigned labels, in a process known as residual region growth, as shown in Fig. 4.8 (analogous to Fig. 3.8). The resultant region growth map at intermediate resolutions levels is combined with an entropy based texture descriptor (M5) utilizing statistical region merging (M6) based on a procedure described in Sections 3.6 and 3.7, to obtain interim segmentations, and the equivalent result at the highest resolution is the final segmentation output of the MAPGSEG algorithm.

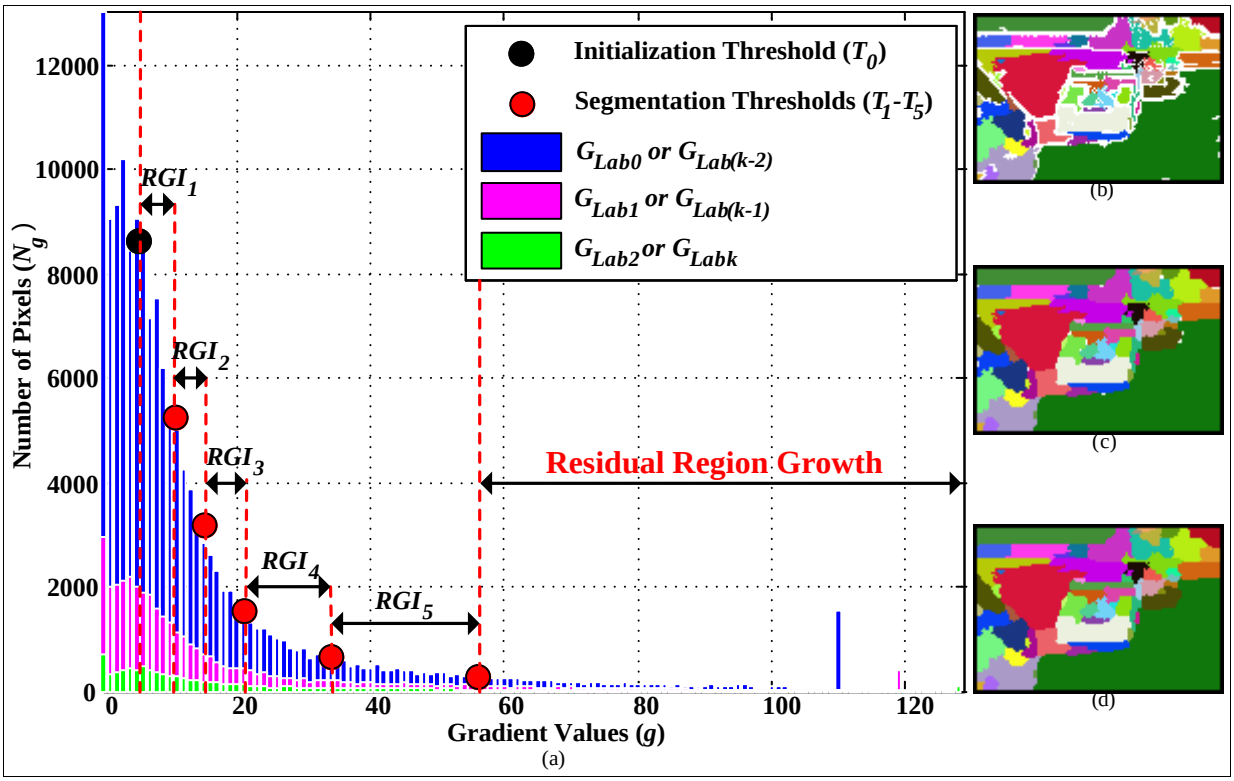

Fig. 4.8: (a) Gradient histogram of the 'Cars' image with adaptively generated segmentation thresholds for $1 \le n \le 5$ and $\Delta g = 10\%$, (b) seed map at the end of five *RGIs*, (c) neighborhood label assignment, (d) iterative morphological label assignment to yield final region growth map.

At an intermediate resolution level, the acquired region growth map is integrated with an entropy based texture descriptor (M5) utilizing a statistical region merging (M6) method, to obtain an interim segmentation output, as shown in Fig. 4.9(a). Furthermore this interim result is exploited by the seed transfer module (M4) to derive seeds that conform well to the semantic distribution of gradient information, of the succeeding dyadic scale image, for facilitating its segmentation with minimal computational costs. Consequently at higher resolution levels, the availability of these pre-segmented regions known as High Confidence Seeds (*HCSs*, see Fig. 4.9(b)), derived from lower level interim segmentations, makes the growth process, required to segment remaining regions (Low Confidence Seeds (*LCSs*)), unwarranted in every *RGI*. Hence the exclusive one-to-one relationship between region growth and seed addition in an arbitrary *RGI*, described previously, is maintained only at the lowest resolution ($L=k$). At higher resolution levels, all *RGI*s are segregated into: 1) intervals that can directly be used for seed addition without region growth, or 2) intervals in which region growth is indispensible before any seed addition can be performed. Thus, if G_{Labj}, $G_{Labj\text{-}HCSs}$, $G_{Labj\text{-}LCSs}$ respectively represent the gradient content of I_{Labj}, *HCSs*, and *LCSs*, at a j^{th} dyadic scale in a k-level decomposition (where $0 \le j < k$ as shown Fig. 9(c) with $j=k\text{-}1=1$, $k=2$), then the decision threshold to perform the aforementioned segregation of various *RGIs*, is the gradient value g which satisfies:

$$N_g(G_{Labj-HCSs}) - N_g(G_{Labj-LCSs}) = 0 \tag{4.2}$$

Here $N_g(G_{Labj-HCSs})$ and $N_g(G_{Labj-LCSs})$ denote the cumulative number of *HCSs* and *LCSs* respectively, with a gradient value *g*. The range of *g* values for which the result of Eq. (4.2) is positive (see region 1 in Fig. 4.9(c)) signifies a gradient range in which the distribution of *HCSs* is larger than *LCSs*, and since most pixels have pre-assigned labels, region growth in this range yields little contribution towards the final segmentation result. Thus *RGIs* in this gradient range are used directly for seed addition without region growth. Conversely, the range of *g* values for which the result of Eq. (4.2) is negative (region 2 in Fig. 4.9(c)) signifies a gradient range largely comprising unsegmented regions, therefore *RGIs* in this gradient range are first engaged in region growth followed by the addition of new seeds. To this effect, region growth at different levels of the input image pyramid 'progresses' towards gradient ranges requiring significant processing, facilitated by seed addition done in a dynamic and distributed framework. Hence we designate this procedure as 'Progressive Region Growth by Distributed Dynamic Seed Addition'.

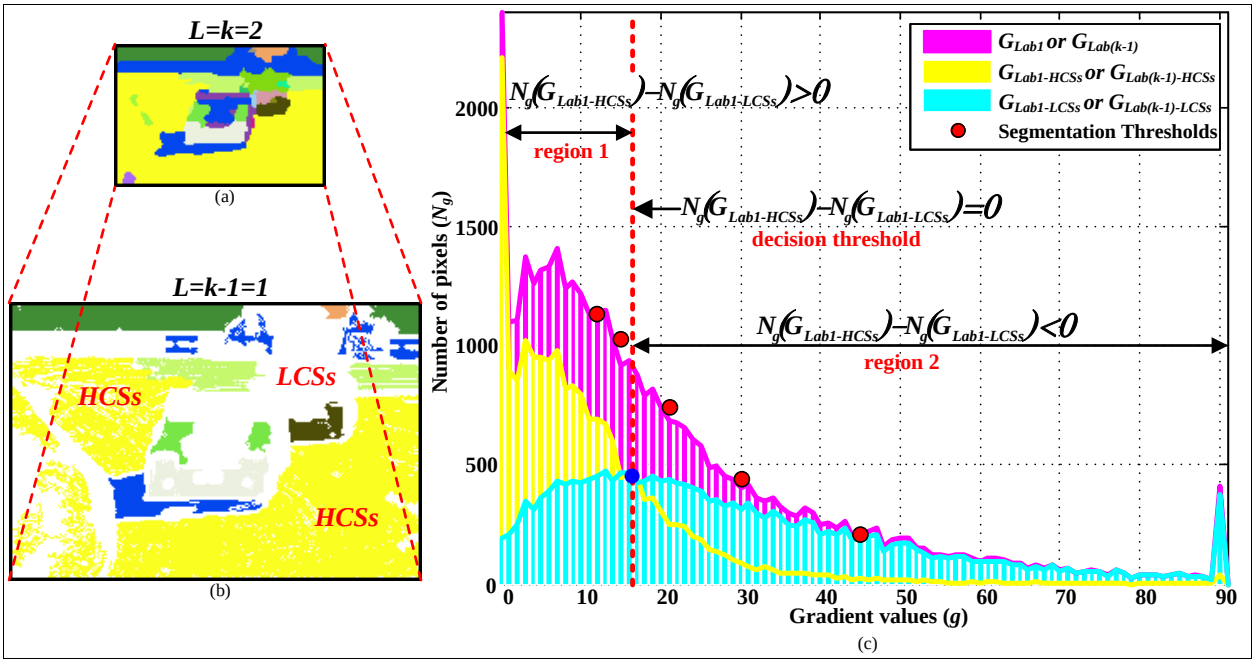

Fig. 4.9: (a) Interim segmentation result at *L=k=2*, obtained at the end of processing in processing in M5 and M6, (b) acquired *HCSs* and *LCSs*, at *L=k-1=1* (j^{th} scale), after the interim segmentation is processed in M4, (c) segregation of RGI utilizing G_{Labj}, $G_{Labj\text{-}HCSs}$, $G_{Labj\text{-}LCSs}$.

4.6. Multi-resolution Seed Transfer

The Multi-resolution Seed Transfer module (M4) is responsible for acquiring *HCSs* from a j^{th} level interim segmentation result at the resolution of $(j\text{-}1)^{th}$ dyadic scale in a *k*-level decomposition (where $0 \le j \le k$ as shown Fig. 4.10(a) with *j=k=2*). Consequently this module can be deemed as an interface for

information transfer from one resolution to another in the MAPGSEG algorithm. An overview of the seed transfer procedure is shown in Fig. 4.10(a).

At an indiscriminate j^{th} level, the acquired interim segmentation result is subjected to a seed map up-conversion process to compute an estimate of it at the resolution of its subsequent dyadic scale for facilitating information transfer, as shown in Fig. 4.10(a). Consequently the interim result (Fig. 4.10(b)) is first up scaled by a factor of two (dyadic scales) by inserting a 'zero' between every pixel, as shown in Fig. 4.10(c), followed by a local neighborhood-based mode filtering discussed previously in Section 3.5. The resultant seed map (see Fig. 4.10(d)), considered an estimate of the j^{th} level segmentation at the subsequent dyadic scale, is utilized to determine seeds that conform well to the semantic distribution of gradient information in the $(j-1)^{th}$ scale, by exploiting a typical characteristic of gradient behavior at different resolutions.

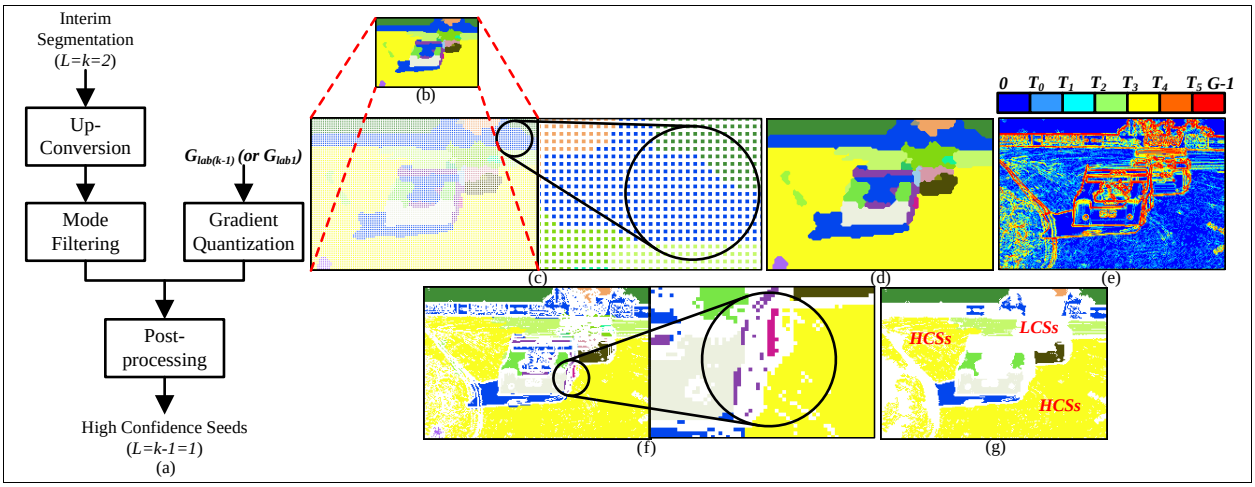

Fig. 4.10: (a) Multi-resolution Seed Transfer module, (b) interim segmentation result (*L=k=2)*, (c) up-conversion (*L=k-1=1)*, (d) mode filtering, (e) gradient quantization, (f) seeds within the two lowest quantization levels, (g) *HCSs* acquired after post processing.

When an image is subjected to wavelet decomposition, regions with uniform or slowly varying gradient can be segmented relatively easily even at coarse resolutions, because there is not much loss in 'image information', due to the decomposition protocol. Conversely, regions with abrupt gradient changes, at coarse resolution levels undergo significant losses in image information and so cannot be segmented with the same ease as done on the full resolution image. Inter-resolution information transfer in the MAPGSEG algorithm exploits this fundamental principle, wherein pixel labels in image areas of uniformly varying gradient at coarse resolutions, and ones in areas of abruptly varying gradient at fine resolutions are respectively retained as *HCSs*. To facilitate this procedure, at an arbitrary resolution the entire gradient gamut is quantized utilizing the initialization threshold (T_0) and segmentation thresholds (T_1 to T_5) yielding a total of seven quantization intervals, as shown in Fig. 4.10(e). Furthermore the

number of gradient quantization intervals to determine *HCSs* for a given scale is based on the total number of resolutions (apart from the smallest resolution since it has no associated a-priori information) to work with, such that there is an approximately even distribution of the quantization intervals among all scales. As an example, for a 2-level decomposition *(k=2)* depicted in Fig. 4.2, pixels labels in the first two quantization intervals are reinstated as *HCSs* from *L=2* to *L=1* (as shown in Fig. 4.10(f)), the next two quantization intervals are employed for information transfer from *L=1* to *L=0*, and consequently region processing at the highest resolution is concentrated only in the final three quantization intervals. Thus information transfer in the MAPGSEG algorithm proceeds from color homogeneous (slowly varying gradient) regions towards regions of color transitions (abrupt variations in gradient) in a manner similar to the way region processing is carried out.

The aforementioned pixel based confidence process, results in numerous small seeds which are isolated as well as mutually adjacent to larger seeds as shown in Fig. 4.10(f). These small regions would result in high computational requirements for processing, due to which reason they are eliminated from the current seed map utilizing connected component analysis based post processing, to yield the *HCSs* map, as shown in Fig. 4.10(g). To this effect, all the remaining unlabeled pixel locations become a part of the *LCSs* map (indicated in Fig. 4.10(g)), which are assigned labels in a fresh run of the algorithm.

4.7. Texture Characterization and Multivariate Region Merging

This section briefly restates the texture modeling (M5) and the region merging procedure (M6) employed in the MAPGSEG algorithm, most part of which have been left unchanged from the GSEG algorithm (described in Sections 3.6 and 3.7 respectively) for region processing at different resolution levels of the input image pyramid. Similar to the GSEG algorithm, the MAPGSEG framework employs an entropy based texture descriptor where-in the entropy of various image segments are calculated to generate a texture map. However in order to achieve computational efficiency by avoiding joint entropy calculation between channels, uniform color quantization of the image data in the CIE L*a*b* color space was employed. The 8-bit encoded color information was uniformly quantized to 6 colors per channel yielding 216 different colors and the result was used for entropy calculation in a 9 by 9 neighborhood. This model of texture is then utilized in the subsequent multivariate region merging process (see Fig. 4.11).

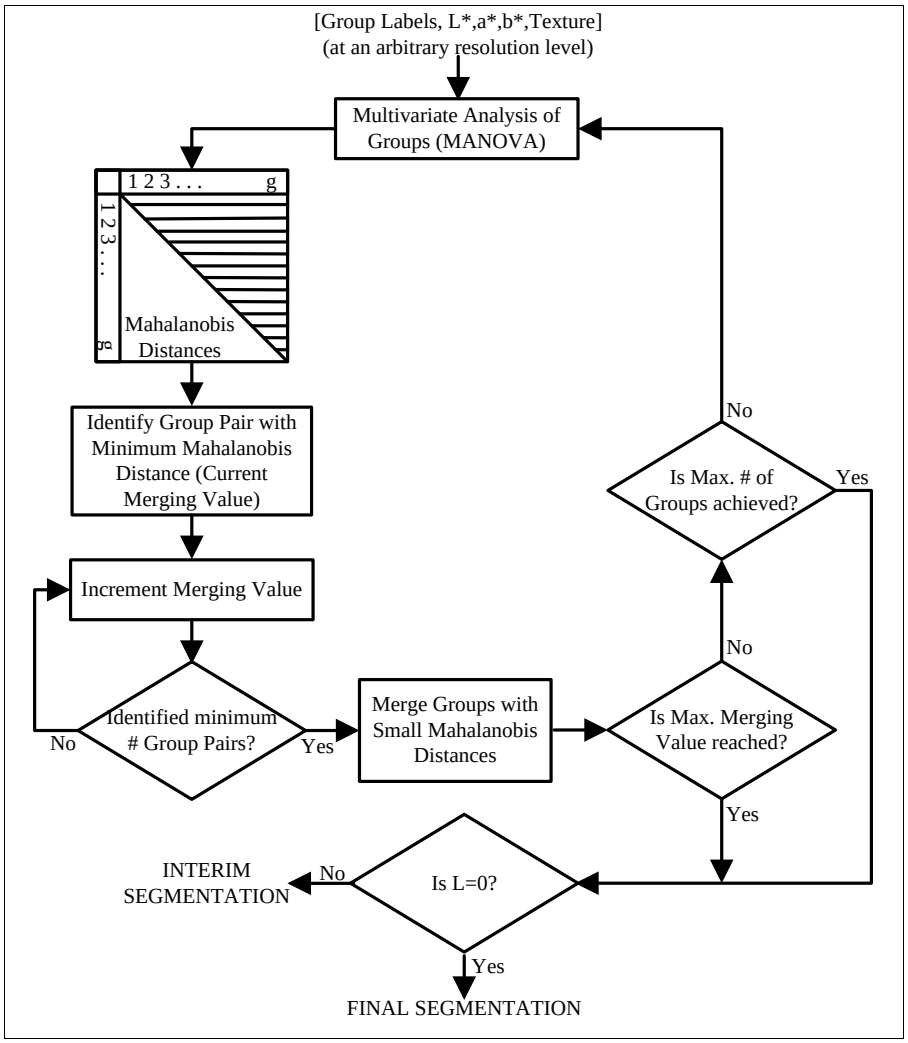

Fig. 4.11: Region merging module (M6) using MANOVA in the MAPGSEG algorithm.

The merging module is utilized to merge regions as deemed necessary, which are over-segmented in the growth procedure due to occlusions and minor texture differences. A multivariate analysis of all independent regions utilizing color and texture is carried out based on the procedure described in Section 3.7. As mentioned previously, the core of a multivariate analysis lies in highlighting the differences between groups that display multiple variables to investigate the possibility that multiple groups are associated with a single factor that facilitates them to be merged together. The multivariate analysis involving the Mahalanobis distance calculation between groups, is carried out on a matrix of dimensions equivalent to the total number of pixels in the image and number of variables (L*, a*, b*, texture) per pixel, for convenient handling of groups. Regions which are most similar are initially found, based on the ones associated minimum Mahalanobis distance (or similarity) value. Moreover, in order to reduce the number of iterations of the merging protocol, the obtained minimum distance value between the two most similar groups is increased until a larger set of candidate group pairs for merging are determined.

Subsequently, the acquired group pairs are merged from the most similar group pair to the least similar pair of the set, to conclude one iteration of the merging process. Subsequently, the Mahalanobis distances for all possible group pairs in the resultant segmentation map are recomputed, and the process is repeated until either a maximum number of acceptable groups (S) is reached or the until the similarity value exceeds a user defined threshold (T), as portrayed in Fig. 4.11. Similar to the GSEG, the two parameters are empirically set at 50 and 2 respectively, although they could be application dependent. The culmination of the region merging process at an arbitrary resolution of the input image pyramid (*L=1, 2,…, k* for a *k*-level decomposition) yields interim segmentations, while at the highest resolution (*L=0*) results in the final segmentation output of the MAPGSEG algorithm, as seen in Fig. 4.11.

4.8 Summary

The MAPGSEG is an efficient method designed for fast segmentation of color images at various resolutions, developed as a multi-resolution extension of the GSEG algorithm with all its novelty. As will be seen in Chapter 7 (Sections 7.1 and 7.2), significant reduction in run time was achieved maintaining benchmark segmentation quality (good balance between quality and speed). The proposed technique is essentially based on the principle that the segmentation results of low-resolution images can be used to efficiently segment their corresponding-high resolution counterparts.

The following chapter discusses a multi-band extension of the GSEG algorithm that proficiently exploits spectral, gradient and textural information, inherent in multi/hyperspectral imagery for use in applications such as 3-D object modeling, large area search, change/target detection etc., to name a few.

Chapter 5

Automatic Spatial Segmentation of Multi/Hyperspectral Imagery by Fusion of Spectral-Gradient-Textural Attributes

Spatial segmentation from a remote sensing perspective, is defined as the process of segregating all pixels in an image cube into spatially meaningful regions (also known as segments or groups) demonstrating homogeneous or nearly homogeneous features. This fundamental image understanding task is employed as a pre-processing step for enhancing the performance of numerous higher-level operations such as 3-D object modeling, large area search, change/target detection, by providing a compact representation of the input scene. The increase in complexity and number of remote sensing applications has necessitated the development of algorithms that effectively perform segmentation in a reasonable timeframe. In this chapter, we propose a new algorithm comprising of five modules (see Fig. 5.1) for multi/hyperspectral image segmentation by extending the GSEG framework (introduced in Chapter 3) to handle multiple bands, based on several spatial and spectral attributes.

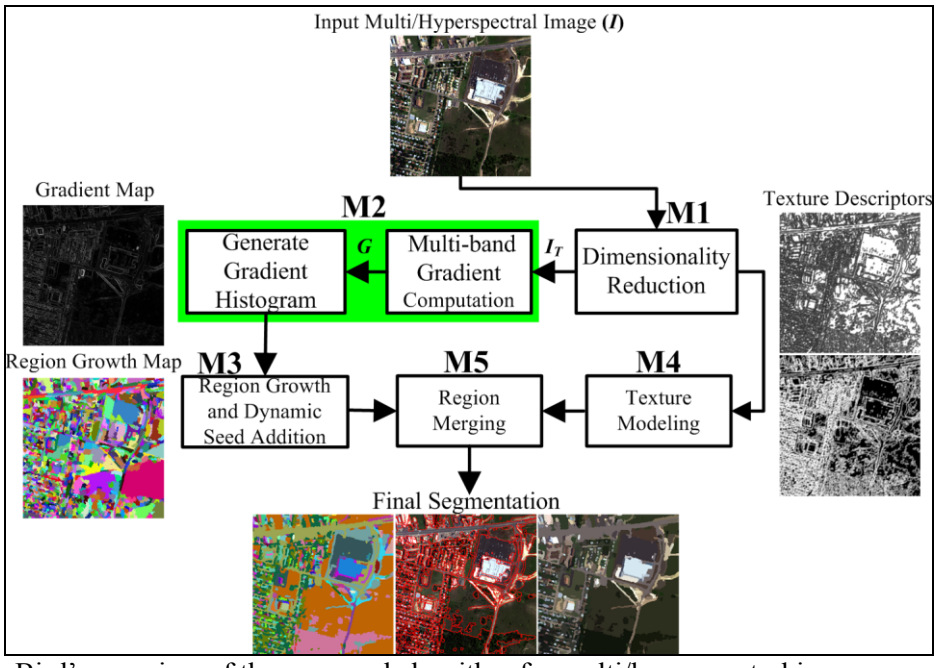

Fig. 5.1: Bird's eye view of the proposed algorithm for multi/hyperspectral image segmentation

The first module (M1) performs dimensionality reduction of the input image cube. The resultant low dimensional dataset is fed to the second module (M2) that uses a vector field gradient estimation approach for deriving gradient/edge information. The third module (M3) is responsible for the formation of an

initial set of regions. The fourth module (M4) is a texture characterization module, which discriminates different textures through features derived from co-occurrence matrices. Finally, the algorithm concludes its operation in the fifth module (M5) which performs a region merging procedure to arrive at the final segmentation result. The following sub-sections describe each of the aforementioned modules in detail.

5.1. Dimensionality Reduction

In this work, dimensionality reduction for multi/hyperspectral imagery is employed as a tool not only to reduce the processing power required for segmentation, but also to improve the quality of the data sets on which it is performed. Initially, the input data (I) is subjected to the Maximum Noise Fraction (MNF) [Green et al., 1988] transformation also known as the Noise Adjusted Principle Component (NAPC) transform [Lee et al., 1990], to facilitate its dimensionality reduction as well as to order various bands by their image quality (from the best to worst). The Maximum Noise Fraction (MNF) transform is a statistical technique that optimizes the Signal-to-Noise Ratio (SNR) of an arbitrary data set. When employing Principle Component Analysis (PCA) (see Appendix A) for data dimensionality reduction, the resultant components do not always show steady decrease in image quality with increasing component number, a scenario that is very common with remotely-sensed data sets. Although the principle components produce a new data space such that the variability in the data is maximum, this new data space does not always capture the maximum variability due to the signal alone. In order to overcome this drawback, the MNF transform is favored over traditional PCA.

The MNF transformation produces new components that show steady decreasing image quality with increasing component number by maximizing the SNR or the noise fraction. If x is a *d*-dimensional multivariate data set where each observation is made of an uncorrelated signal and noise component, then x can be represented as:

$$x = \begin{bmatrix} S(x_1) + N(x_1) \\ \vdots \\ S(x_d) + N(x_d) \end{bmatrix} \tag{5.1}$$

Furthermore if the noise and signal are assumed to be uncorrelated, then the covariance of such a data set can be computed as the sum of the individual signal and noise covariance matrices:

$$\Sigma_x = \Sigma_S + \Sigma_N \tag{5.2}$$

where, the noise covariance matrix required in Eq. (5.2) is estimated as:

$$\Sigma_N \cong \Sigma_\Delta = \Sigma_{\frac{1}{2}(x - x_\Delta)} \tag{5.3}$$

Here $x - x_\Delta$ is an image formed by subtracting a slightly offset version of the original data from itself. In general since the signal component exhibits high spatial correlation within a band, while the noise

component of the signal shows little spatial variation, Eq. (5.3) will result removing the underlying signal value while leaving the statistics of the noise unaffected.

Thus, the noise fraction for an arbitrary i^{th} band is defined as the ratio of the total noise variance, to the total variance for that band:

$$NF(x_i) = \frac{\sigma^2_{N(x_i)}}{\sigma^2_{(x_i)}} \tag{5.4}$$

To this effect, the MNF transformation chooses a linear transformation:

$$y = Gx \tag{5.5}$$

such that the noise fraction for y_i is maximum among all linear transformations orthogonal to y_j for all $j > i$. Moreover, the individual transforms that make up the matrix G are defined as the eigenvectors of the matrix $\Sigma_N \Sigma_x^{-1}$ and the corresponding eigenvalues are the individual noise fractions ($NF(x_i)$). In performing the MNF transformation the noise fractions and corresponding noise fraction transforms are ordered such that:

$$NF(x_1) \geq NF(x_2) \geq NF(x_3) \geq \cdots \geq NF(x_d) \tag{5.6}$$

Since the signal and noise are assumed to be uncorrelated, the orthogonalization of x also orthogonalizes the signal $S(x)$, and noise $N(x)$, terms. Having sorted the noise fractions and corresponding noise fraction transforms according to Eq. (5.6), noise removal occurs by spatial filtering the nosiest components of y and then back transforming the data set to the original domain, which can be represented by the expression below:

$$x' = G^{-1} y_{filtered} \tag{5.7}$$

Smoothing in this transformed space allows for much stronger noise removal than simply smoothing in the original space where the actual spatial information would also by subjected to smoothing, resulting in the degradation of the original data. The data set x', is then subjected to the PCA algorithm (discussed in Appendix A), which maximizes the variance along the each component axis where the variance is now exclusively due to the signal of concern, resulting in a final transformed image (I_T) with reduced dimensions.

5.2. Multiband Gradient Detection for Multi/Hyperspectral Imagery

The dimensionality reduction step is succeeded by a gradient detection scheme that is employed on the transformed image (I_T), to ensure that the regions formed are consistent with the spectral intensity variations across the scene. Abstraction of edge information from multi/hyperspectral data has often been

a challenging task. We propose an efficient approach for achieving highly localized edges in remotely sensed imagery (see Fig. 5.2 for a block diagram).

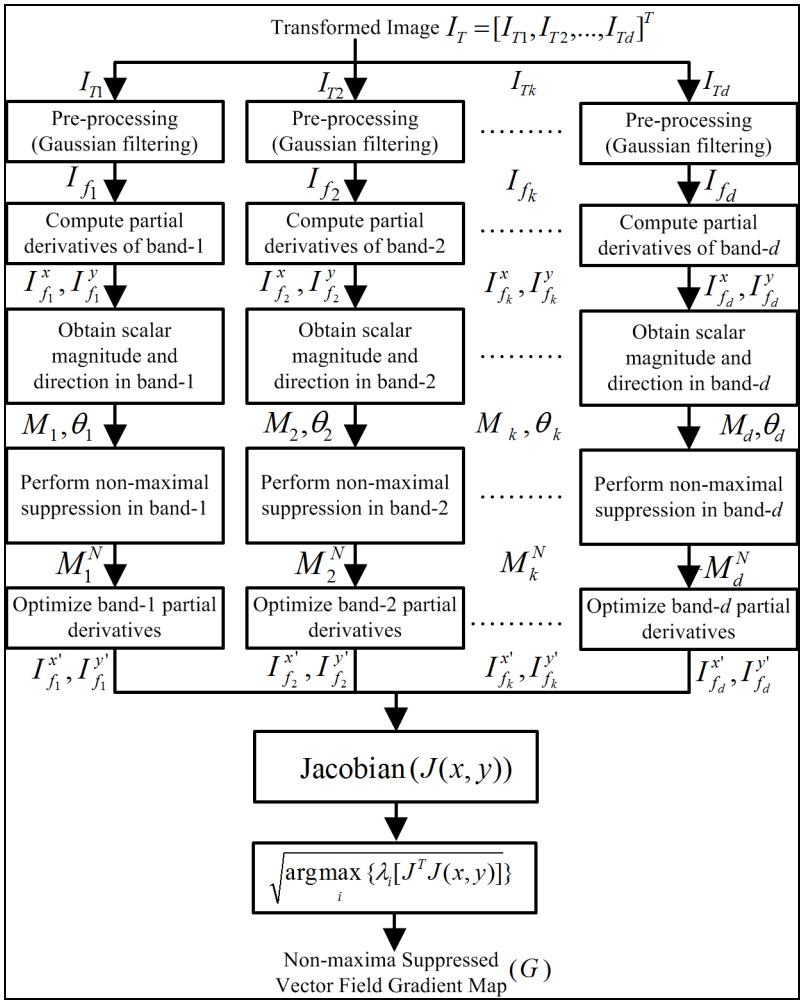

Fig. 5.2: Proposed gradient computation scheme

Our multiband edge detection algorithm is based on the notion that the partial derivatives of individual image bands used in a typical vector field gradient estimation method often yields thick edges with spurious responses around true edge points. Consequently, we believe that optimizing the partial derivatives of various bands to only comprise of contributions towards their associated local scalar gradient maxima, before employing them in a vector gradient calculation, can yield highly localized responses in the resultant edge map.

In what follows, let $I_T = [I_{T1}, I_{T2}, ..., I_{Td}]^T$, where I_{Tk} denotes the k^{th} band while d is the total number of bands. Furthermore, let x, y represent the spatial co-ordinates such that $1 \le x \le K$, and $1 \le y \le L$. Our method for gradient detection begins by pre-processing each individual band of the image ($I_{Tk}, k \in \{1,2,...,d\}$) using a Gaussian filtering scheme (with a standard deviation

parameter $\sigma = 1$) tuned to extract edges at a specific scale as well as minimize the effects of noise. This filtering process can be mathematically represented by a 2-D convolution as:

$$I_{fk}(x,y) = I_{Tk}(x,y) * G_a(x,y) = I_{Tk}(x,y) * \frac{1}{2\pi\sigma} e^{-(x^2+y^2/2\sigma^2)} \tag{5.8}$$

where, I_{fk} is the filtered version of I_{Tk}, and $G_a(x,y)$ symbolizes a 2-D Gaussian function. Subsequently, the partial derivatives (I_{fk}^x, I_{fk}^y) of I_{fk} with respect to the spatial image co-ordinates x and y, are computed using the widely-known Prewitt operators [Prewitt, 1970], given as:

$$I_{fk}^x(x,y) = \frac{\partial I_{fk}(x,y)}{\partial x} = \{I_{fk}(x-1,y+1) + I_{fk}(x,y+1) + I_{fk}(x+1,y+1)\} - \{I_{fk}(x-1,y-1) + I_{fk}(x,y-1) + I_{fk}(x+1,y-1)\} \tag{5.9}$$

$$I_{fk}^y(x,y) = \frac{\partial I_{fk}(x,y)}{\partial y} = \{I_{fk}(x+1,y-1) + I_{fk}(x+1,y) + I_{fk}(x+1,y+1)\} - \{I_{fk}(x-1,y-1) + I_{fk}(x-1,y) + I_{fk}(x-1,y+1)\} \tag{5.10}$$

It is imperative to observe that any set of operators to calculate the derivatives can be used. The resultant derivatives are utilized to determine the scalar gradient magnitude (M_k) and direction (θ_k) associated with each band, given as:

$$M_k(x,y) = \sqrt{\{I_{fk}^x(x,y)\}^2 + \{I_{fk}^y(x,y)\}^2} \tag{5.11}$$

$$\theta_k(x,y) = \tan^{-1}(I_{fk}^y(x,y) / I_{fk}^x(x,y)) \tag{5.12}$$

where $\theta_k(x,y)$ in Eq. (5.12) is orthogonal to the direction of the edge at (x,y). Next, the scalar magnitude and direction of the gradient resulting from each image band is individually subjected to a non-maximal suppression scheme [Canny, 1986], as depicted in Fig. 5.2.

Non-maximal suppression is a two-step procedure utilized to suppress spurious responses around true edge points (related to local maxima) of a scalar field, uncovering highly localized edge components. First, the obtained gradient direction ($\theta_k(x,y)$) is quantized to one of four discrete orientations $\theta_1 = 0^o$, $\theta_2 = 45^o$, $\theta_3 = 90^o$, $\theta_4 = 135^o$ representing vertical, horizontal and diagonal directions respectively. Following this, suppression of spurious edge responses is accomplished by finding the discrete orientation ($\theta_o, o \in \{1,2,3,4\}$) closest to $\theta_k(x,y)$ in a 3x3 neighborhood (β) encompassing every pixel (x,y), and eliminating its corresponding gradient magnitude $M_k(x,y)$ if and only if its value is smaller than one of its two abutting neighbors ($\{\beta_1, \beta_2\} \in \beta$) along θ_o [Canny, 1986, Gonzalez and Woods, 2008]. This operation is mathematically represented as:

$$M_k^N(x,y) = \begin{cases} 0 \;\; if \;\; M_k(x,y) < \beta_1 \; or \; M_k(x,y) < \beta_2 \\ M_k(x,y) \qquad\qquad\qquad otherwise \end{cases} \tag{5.13}$$

In Eq. (5.13), M_k^N is the non-maximally suppressed equivalent of M_k exclusively containing gradient magnitudes associated only with true edge points (or local gradient maxima). Having acquired M_k^N for an arbitrary k^{th} image band, the partial derivatives I_{fk}^x, I_{fk}^y utilized to compute M_k are optimized in accordance with M_k^N. This is done to ensure that these derivatives that are eventually utilized in a vector field gradient calculation comprise of values only contributing to gradient magnitudes of true edge points while contributions to spurious responses are suppressed. More specifically, the values of $I_{fk}^x(x,y)$ and $I_{fk}^y(x,y)$ are retained only if the scalar gradient magnitude at (x,y) before and after non-maximal suppression remains unaltered, implying a true edge point at which no suppression has taken place. Conversely, the values of $I_{fk}^x(x,y)$ and $I_{fk}^y(x,y)$ are suppressed when the scalar gradient magnitude before and after non-maximal suppression has changed implying a spurious edge point at which suppression has taken place. This optimization can be represented as:

$$I_{fk}^{x'}(x,y) = \frac{\partial I_{fk}'(x,y)}{\partial x} = \begin{cases} 0 & if\ M_k^N(x,y) \neq M_k(x,y) \\ I_{fk}^x(x,y) & if\ M_k^N(x,y) = M_k(x,y) \end{cases} \tag{5.14}$$

$$I_{fk}^{y'}(x,y) = \frac{\partial I_{fk}'(x,y)}{\partial y} = \begin{cases} 0 & if\ M_k^N(x,y) \neq M_k(x,y) \\ I_{fk}^y(x,y) & if\ M_k^N(x,y) = M_k(x,y) \end{cases} \tag{5.15}$$

Eqs. (5.14), (5.15) can also be supported by the fact given that the scalar gradient magnitude at a pixel is zero (after performing non-maximal suppression), the only possible real solution to Eq. (5.11) is when both the partial derivatives are also zero. The optimized derivatives acquired using Eqs. (5.14), (5.15) across all bands are subsequently employed in a vector field gradient calculation method [Lee and Cok, 1991] to ensure information from all bands are taken into consideration while generating the final edge map. To facilitate this, a spatial gradient matrix (J) of all first order partial derivatives (also known as the Jacobian matrix), at every pixel location (x, y) is constructed as:

$$J(x,y) = \begin{bmatrix} I_{f1}^{x'}(x,y) & I_{f1}^{y'}(x,y) \\ I_{f2}^{x'}(x,y) & I_{f2}^{y'}(x,y) \\ \vdots & \vdots \\ I_{fd}^{x'}(x,y) & I_{fd}^{y'}(x,y) \end{bmatrix} = \begin{bmatrix} \frac{\partial I_{f1}'(x,y)}{\partial x} & \frac{\partial I_{f1}'(x,y)}{\partial y} \\ \frac{\partial I_{f2}'(x,y)}{\partial x} & \frac{\partial I_{f2}'(x,y)}{\partial y} \\ \vdots & \vdots \\ \frac{\partial I_{fd}'(x,y)}{\partial x} & \frac{\partial I_{fd}'(x,y)}{\partial y} \end{bmatrix} \tag{5.16}$$

The magnitude of the d -dimensional vector field at location (x, y) is acquired by first constructing an inner product matrix $J^T J(x,y)$, given by:

$$J^T J(x,y) = \begin{bmatrix} a(x,y) & c(x,y) \\ c(x,y) & b(x,y) \end{bmatrix} \tag{5.17}$$

where,

$$a(x,y) = \sum_{k=1}^{d} \left(\frac{\partial I'_{fk}(x,y)}{\partial x} \right)^2 \tag{5.18}$$

$$b(x,y) = \sum_{k=1}^{d} \left(\frac{\partial I'_{fk}(x,y)}{\partial y} \right)^2 \tag{5.19}$$

$$c(x,y) = \sum_{k=1}^{d} \left(\frac{\partial I'_{fk}(x,y)}{\partial x} \right) \left(\frac{\partial I'_{fk}(x,y)}{\partial y} \right) \tag{5.20}$$

Then, the gradient magnitude (G) is computed by finding the square root of the largest Eigenvalue of $J^T J(x,y)$, given by:

$$G(x,y) = \sqrt{\arg\max_i \{\lambda_i [J^T J(x,y)]\}} \tag{5.21}$$

In Eq. (5.21) it can be recognized that $J^T J(x,y)$ has two Eigenvalues ($\lambda_i, i = 1,2$) calculated specifically at (x,y), while the notation $\lambda_i[J^T J(x,y)]$ represents the i^{th} Eigenvalue of $J^T J(x,y)$. Also, since $J^T J(x,y)$ is a real, symmetric and positive semi-definite 2x2 matrix, the close form solution of its largest Eigenvalue can be given as [Lee and Cok, 1991]:

$$\arg\max_i \{\lambda_i [J^T J(x,y)]\} = \frac{1}{2}\{a(x,y) + b(x,y) + \sqrt{[a(x,y) + b(x,y)]^2 - 4[a(x,y)b(x,y) - c(x,y)^2]}\} \tag{5.22}$$

In Fig. 5.3(a), a hyperspectral 'Urban' scene of dimensions 307 X 307 X 210 obtained from the Hyperspectral Digital Imagery Collection Experiment (HYDICE) sensor [Mitchell, 1995] is displayed as an RGB color composite using bands 49, 35 and 18. Figs. 5.3(b), 5.3(c) reveal edge information derived using the first 8 MNF bands (>99% of data variability), from our implementation of a vector field gradient detection algorithm described in [Lee and Cok, 1991] as well as the previously described methodology. Observation of the aforestated results substantiates the advantage of our approach in deriving edge information. More specifically, the uniqueness of our edge detection protocol lies in the fact that it leverages the non-maximal suppression scheme to enhance (or localize) edges of a vector field, as opposed to its conventional use for enhancing edge outcomes of a scalar field. Having established the gradient content (G) of the transformed image I_T a histogram of these gradient values is generated, to be employed in subsequent stages of processing, as depicted in Fig. 5.1.

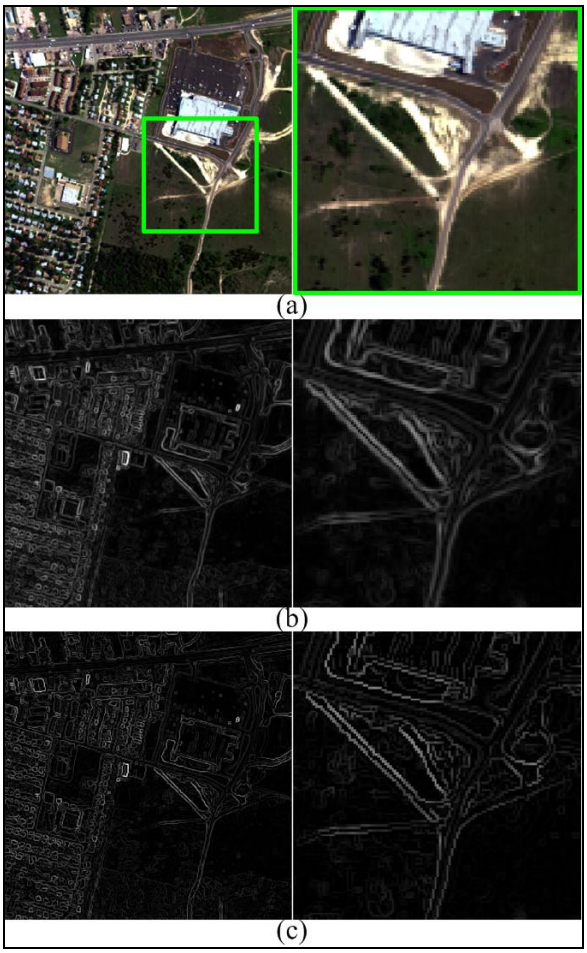

Fig. 5.3: (a) HYDICE 'Urban' scene (bands 49, 35, and 18), (b) Result using [Lee and Cok, 1991], and (c) Result using proposed gradient computation scheme.

5.3. Initial Clustering and Region Growing

This section re-capitulates the region formation process (M3) described in Fig. 5.3, a significant portion of which has been used unaltered from the GSEG algorithm (described in Sections 3.4 and 3.5 respectively). Region formation, in the proposed algorithm involves two stages namely: initial clustering and region growing, as shown in Fig. 5.5. The initial clustering step triggers region compilation by identifying pixel-congregations in "flat areas" possessing gradually varying spectral signatures corresponding to small gradient magnitudes within them. Such an ensemble of homogeneous pixel-groupings are determined by choosing an initialization threshold (g_{init}) confined to the lower end of the gradient histogram. One way of accomplishing this is by choosing g_{init} as a small percentage (α) of the maximum value ($G_{\max}$) of the gradient histogram, given by $g_{init} = \alpha G_{\max}$. Next, pixels in G that satisfy the condition $G \le g_{init}$ are subjected to connected component analysis [Haralick and Shapiro, 1992], to yield a map comprised of several uniquely labeled pixel-clusters called Parent Seeds (PSs), as shown in

Fig. 5.4(a). It is important to note that the seeds in Fig. 5.4 are randomly color coded for distinction. Furthermore, to evade generation of numerous small seeds in homogeneous areas, a size criterion is utilized to restrict the spatial extent of all *PSs* (see Fig. 5.4(b)).

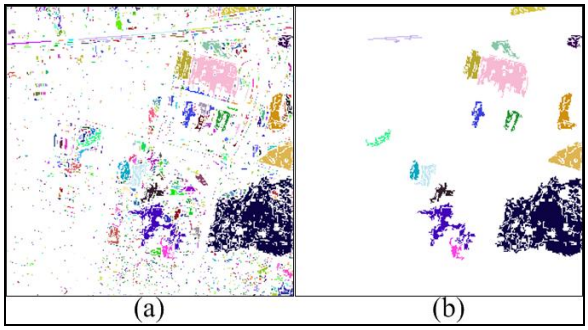

Fig. 5.4: Initial clustering: (a) Initial parent seeds (*PSs*) map using g_{init}=2.5 %{G_{max}}, and (b) Spatially constrained *PSs* map using 20**MSS*=100pixels.

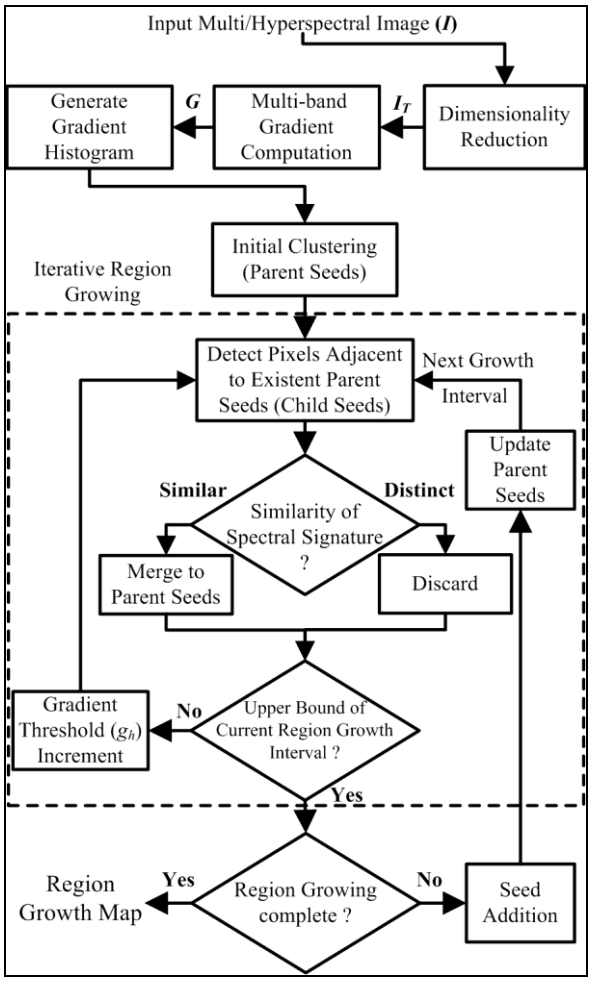

Fig. 5.5: Flowchart of the proposed region growing method

In the proposed algorithm, spatial constraints required at various stages of region formation are chosen relative to a Minimum Seed Size (*MSS*) criterion that defines the smallest possible size of an individual

seed. The *MSS* measure is set (empirically at 5 pixels) such that it is small enough to ensure that reasonably fine details in the image are captured as the algorithm progresses. Therefore, to ascertain proper region formation, the *PSs* generated in the initial clustering stage aimed at capturing low gradient regions are restricted to a large spatial constraint of ~20 times *MSS*, while this criterion is gradually reduced for addition of new seeds at subsequent higher gradient densities at which finer image details manifest themselves. Generation of *PSs* concludes the initial clustering stage, which is succeeded by an iterative region growing stage.

Traditionally, region growing is defined as a process which starts from a single pixel or a set of individual pixels each termed as a seed point and based on a specific similarity/ homogeneity rule that iteratively accumulates pixels around it. The growth of a region stops when no more pixels satisfying the homogeneity principle are found, thus facilitating formation of spatially compact regions. Most conventional region growing approaches suffer from some or all of the following short comings: (a) they require manual selection of seed points, (b) they perform pixel by pixel agglomerations, (c) they have a sequential design (handle one region at a time), (d) they tend to have a processing order dependency, and (e) they are expensive in terms of computation and memory requirements. Conversely, we propose a region growing approach that simultaneously lays emphasis on the homogeneous and heterogeneous characteristics of pixels in an image, using spectral and gradient information. The process commences at pixel sites with small gradient magnitudes (gradually varying spectral intensity) and terminates at locations with large gradient magnitudes (abrupt spectral intensity variations) with no processing-order dependency. The growth approach is integrated with an active seed addition mechanism that automatically identifies new seeds in un-segmented areas at distinct stages of processing, consequently accommodating parallel growth of multiple adjacent and/or non-adjacent image regions. An overview of the proposed region growing algorithm is displayed in Fig. 5.5.

The region growing process proceeds to grow *PSs* by first identifying the next gradient magnitude (g_h), higher than g_{init} in the histogram of G. Following this, pixels locations not part of *PSs* and whose gradient values are less than or equal to g_h undergo connected component labeling to yield a map comprising of a new seeds (Fig. 5.6(a)), designated as Child Seeds (*CSs*). At this stage, only the *CSs* adjoining the existent *PSs* are overseen. These are identified by determining the ones that possess pixels intersecting with the peripheral boundaries of *PSs*. Parent Seed boundaries ($PS_{boundaries}$) are morphologically extracted (Fig. 5.6(b)) as a binary map by subtracting them from their dilated equivalents, mathematically represented as:

$$PS(x,y)_{boundaries} = \begin{cases} 1 \forall\ (x,y)\ where\ \{(PSs \oplus \psi) - PSs\} > 0 \\ 0 \forall\ (x,y)\ where\ \{(PSs \oplus \psi) - PSs\} = 0 \end{cases} \tag{5.23}$$

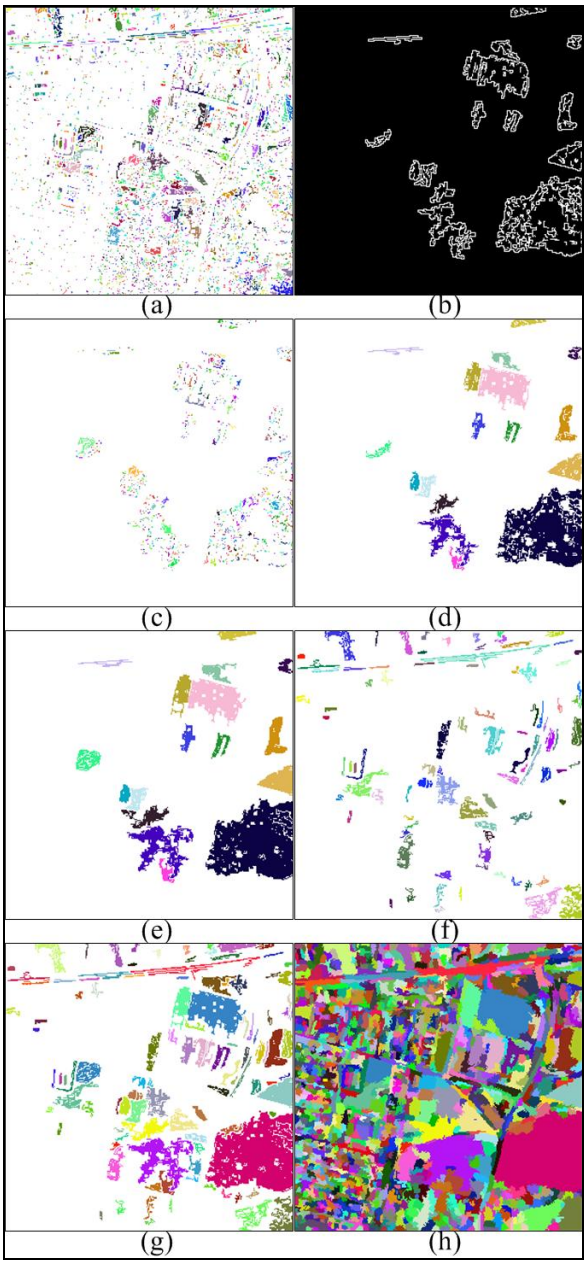

Fig. 5.6: (a) Child seeds (*CSs*) map, (b) Parent seed boundaries ($PS_{boundaries}$), (c) Adjacent *CSs*, (d) Region map obtained after parent-child similarities have been processed, (e) Region map obtained at the end of a few iterations of the growth process, (f) New seeds non-adjacent to existent *PSs*, (g) *PSs* for next stage of region growth, and (h) Final region growth map.

In Eq. (5.23), $\oplus$ represents a morphological dilation operation while ψ denotes a 3x3 uniform (all 1's) structuring element. On conclusion of the aforementioned computation, the labels of *CSs* bearing adjacency with *PSs* are obtained (Fig. 5.6(c)) by performing a pixel-wise multiplication of the *CSs* map and the binary $PS_{boundaries}$ map. Furthermore, to ensure that the growth process is computationally efficient, all adjacent *CSs* whose sizes are smaller than the *MSS* criterion are directly assigned their associated

parent labels, while the larger ones are spectrally compared to their parents. Parent-child spectral similarity is assessed by computing the normalized Euclidean distance ($D_{NE}^{PSs-CSs}$), between their mean spectral signatures ($\vec{m}_{PSs} = [p_1,...,p_d]^T$) and ($\vec{m}_{CSs} = [c_1,...,c_d]^T$), denoted as:

$$D_{NE}^{PSs-CSs} = \sqrt{\sum_{k=1}^{d} (p_k - c_k)^2 / \sigma_k^2} \tag{5.24}$$

where σ_k^2 is the spectral intensity variance in the k^{th} band. Thus, the adjacent *CSs* (larger than *MSS*) whose mean spectral signatures are found in close proximity to their parents are merged with them, while the rest are discarded, as shown in Fig. 5.5. Specifically, the aforestated parent-child closeness of spectral information is deduced by evaluating whether $D_{NE}^{PSs-CSs}$ values are less than a maximum allowed distance D_{NE}^{max} set equal to 1. This culminates in a single iteration of region growing at a gradient threshold g_h, resulting in the gradual growth of *PSs* (see Fig. 5.6(d) relative to Fig. 5.4(b)). Moreover, as the aforementioned procedure progresses iteratively into higher gradient densities effecting continued and simultaneous growth of all *PSs*, image areas that are not adjacent to them remain un-segmented in the region map, as portrayed in Fig. 5.6(e). Therefore, at specific processing intervals our growth mechanism is interleaved with an active seed addition step to facilitate addition of new seeds in areas that are non-adjacent to the existent *PSs*. These intervals, as discussed in previous chapters are termed region growth intervals, and are derived by dividing the entire gradient histogram into M sections demarcated using $M-1$ gradient values, such that each section spans approximately equal image area (or number of pixels). Fig. 5.7 shows the gradient histogram of a HYDICE image previously displayed in Fig. 5.3(a), fragmented into ten region growth intervals (labeled in green and distinguished using red lines) each spanning equal (~10%) image area. Note that the aforestated choice of the number of intervals was found to optimally regulate (i.e. not being too many or too few) the number of seeds being added to the region map, while accounting for the exponential decay in the gradient histogram. Thus, on conclusion of the growth process in an arbitrary interval (when g_h is equal to the upper bound), pixel-clusters in areas that are not adjacent to existing seeds and larger than the *MSS* criterion (Fig. 5.6(f)) are added to the region map. The resulting agglomeration of seeds becomes the *PSs* map for the next interval of region growing, as seen in Figs. 5.5 and 5.6(g). This interlaced mechanism involving region growing and active seed addition culminates when the gradient threshold equivalent to the highest magnitude $G_{\max}$ is processed, to arrive at the final region growth map (Fig. 5.6(h)).

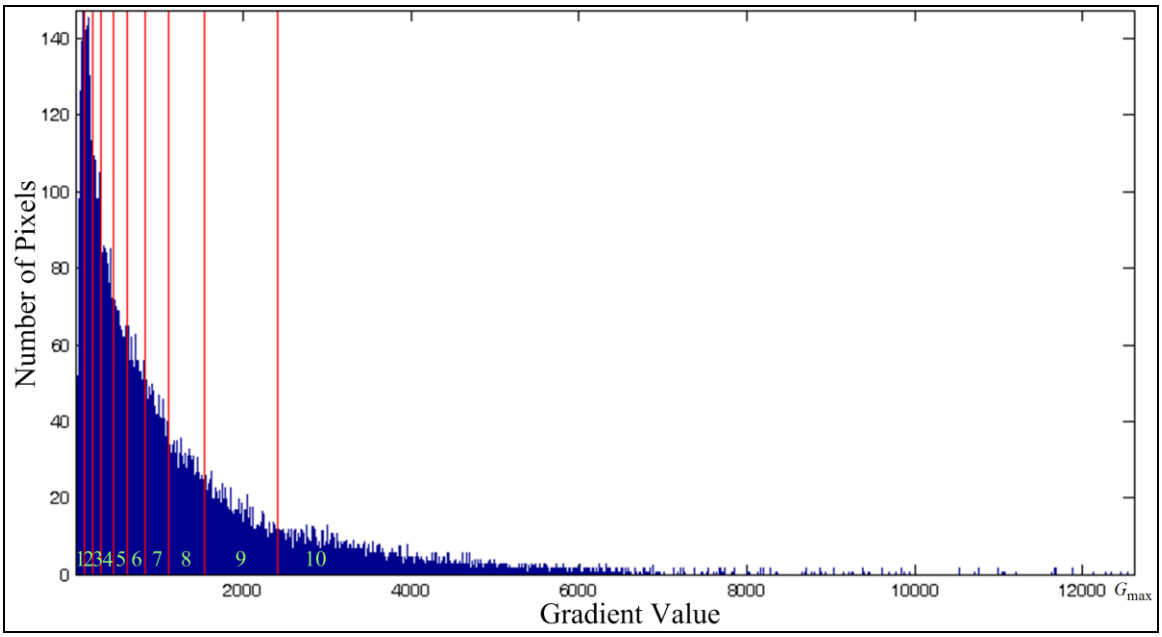

Fig. 5.7: Gradient histogram of the HYDICE image shown in Fig. 5.3(a) divided into ten region growth intervals each spanning equal image area.

5.4. Co-occurrence Features based Texture Modeling

Effective segmentation in the presence of texture is an exigent task. Thus, the proposed algorithm was equipped with a texture extraction component (M4 in Fig. 5.1), which characterizes different textures via descriptors derived from co-occurrence matrices [Haralick et al., 1973]. These matrices help construct textural attributes based on computing the gray-tone spatial dependence frequencies among pixels as a function of the relative orientation (ϕ) and distance (δ) between them. For a pair of pixels located at (k,l) and (u,v) respectively, and associated with gray levels i, j, the un-normalized gray-tone spatial dependence matrices (P) for all possible values of ϕ (typically quantized to 45^o intervals: $0^o, 45^o, 90^o$, and 135^0) and a single distance parameter ($\delta = 1$), are defined as:

$$P(i,j,\delta=1,\phi=0^o) = \#\{(k,l),(u,v)\} \in \beta \mid (k-u=0, |l-v|=\delta), \beta(k,l)=i, \beta(u,v)=j\} \quad (5.25)$$

$$P(i,j,\delta=1,\phi=90^o) = \#\{(k,l),(u,v)\} \in \beta \mid (|k-u|=\delta, l-v=0), \beta(k,l)=i, \beta(u,v)=j\} \quad (5.26)$$

$$P(i,j,\delta=1,\phi=45^o) = \#\{(k,l),(u,v)\} \in \beta \mid (k-u=\delta, l-v=-\delta) \text{ or } (k-u=-\delta, l-v=\delta), \beta(k,l)=i, \beta(u,v)=j\} \quad (5.27)$$

$$P(i,j,\delta=1,\phi=135^o) = \#\{(k,l),(u,v)\} \in \beta \mid (k-u=\delta, l-v=\delta) \text{ or } (k-u=-\delta, l-v=-\delta), \beta(k,l)=i, \beta(u,v)=j\} \quad (5.28)$$

where '#' denotes the number of elements in each set. In Fig. 5.8(a) shown is a sample 4x4 image matrix consisting of four gray levels (0, 1, 2 and 3). Displayed in Fig. 5.8 (b) is the un-normalized gray tone spatial dependence frequency matrix (P) computed for $\phi = 0^o$ and $\delta = 1$. The element in the position (1, 2) of P is the cumulative number of times the two gray level values 0 and 1 occurred in the horizontal

direction (since $\phi = 0^o$) at spacing of one pixel (since $\delta = 1$). From result in Fig. 5.4 (b) it can be discerned that P is a symmetric matrix whose dimensions are dependent on the number of gray levels in image.

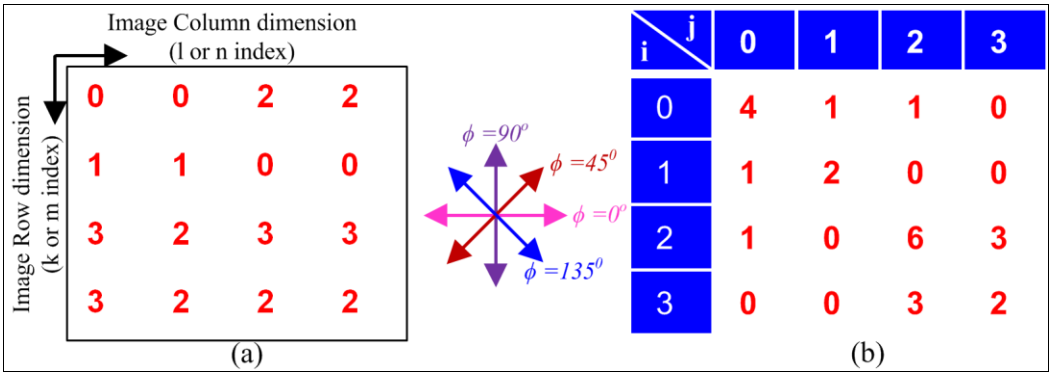

Fig. 5.8: (a) sample 4x4 image matrix, (b) computed co-occurrence matrix or gray tone spatial dependence frequency matrix (P) calculation using $\phi = 0^o$ and $\delta = 1$.

Using normalized versions of the Gray-Tone Spatial-Dependence Matrices by dividing each element of P with the total number of pixel pairs (R) in the image (for the example in Fig. 5.8 R =24 which can be obtained by adding all the entries in P), Haralick et al. defined 14 features which could be potentially used as texture descriptors. However in this work, five of the most extensively used descriptors for multi/hyperspectral data, namely angular second moment (f_{ASM}), contrast (f_{CON}), variance (f_{VAR}), inverse difference moment (f_{IDM}) and entropy (f_{ENT}) defined in Eqs. (5.35) - (5.39), were used.

$$f_{ASM} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} \{P(i,j)/R\} \tag{5.29}$$

$$f_{CON} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i-j)^2 \{P(i,j)/R\} \tag{5.30}$$

$$f_{VAR} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} (i-\mu)^2 \{P(i,j)/R\} \tag{5.31}$$

$$f_{IDM} = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} [1/\{1+(i-j)^2\}] \{P(i,j)/R\} \tag{5.32}$$

$$f_{ENT} = -\sum_{i=1}^{N_g} \sum_{j=1}^{N_g} P(i,j) \log[P(i,j)/R] \tag{5.33}$$

Each of the features defined in Eqs. (5.29) - (5.33) provide certain information about an image. Angular second moment and inverse difference moment are two measures of homogeneity that possess high values when the scene has only a few gray levels. On the contrary variance is a measure of heterogeneity that attains high values for rapidly varying gray levels in the scene. Contrast is a measure of local intensity variations present in the image and similar to variance it acquires high values for large variations of gray levels. Lastly, as mentioned in Section 3.6 entropy represents the degree of randomness present in the

image, and consequently inhomogeneous scenes typically have high entropy while homogeneous ones have low entropy values.

In this work, five texture maps were generated by computing the aforementioned feature values from a co-occurrence matrix extracted from a local 3x3 neighborhood (β) around every image pixel. Moreover, to reduce computations, the features were derived only from data in the first MNF band having the highest image quality, using a single arbitrarily selected orientation ($\phi = 45^o$) and a distance parameter ($\delta = 1$). Note that the aforesaid band selection and choice of ϕ and δ parameters does not intercede with the objective of modeling textural content, since our intention was to achieve its coarse representation from the image rather than a precise rotationally invariant depiction (using multiple values of ϕ and δ across all bands) that necessitates a significant increase in computational load. Moreover in the case of multi/hyperspectral data that usually contain 8-12 bits of information per pixel the resultant co-occurrence matrices generated can be quite large and sparsely populated, due to which the first MNF band was quantized to fewer gray levels before being employed for texture characterization [Schott, 2007].

5.5 Multivariate Region Merging

As mentioned previously, the region map formed at the culmination of the growing process yields an over-segmented result, achieved primarily based on the spectral signature similarity of pixels traversing various gradient densities, as well as due to abundant presence of texture disparities in the scene. Consequently, this region map is optimized by fusing spectral intensity (from all MNF bands), with texture information in a unique merging procedure driven by a statistical routine called the Multivariate Analysis of Variance (MANOVA), described in detail in Section 3.7. The essence of this method is to formulate an alternate canonical variate space where-in the Euclidean distance between group means (of the variates involved) is equivalent to Mahalanobis distance between them in the original data space.

On initiation of the merging process (M5 in Fig. 5.3), the region-pair with the minimum distance value, signifying the maximum similarity is found. However, instead of fusing only a single group pair per merging iteration, we adopt a "greedy" algorithm for computational efficiency, where the distance between the two most similar groups is gradually increased until a larger set of "highly similar" groups pairs are obtained, before merging is performed. In this subset of similar group pairs, regions are fused with each other from the most similar pair to the least similar one, which culminates a single merging-iteration. Next, the distances are recomputed for all possible group pairs in the new/refined set of multivariate populations. This process is repeated hierarchically until the distance value exceeds a user-defined threshold (T =1) or the desired number of segments (S set based on complexity of the image) are

obtained, which yields the final segmentation outcome. The first criterion ensures the acquisition of a functional number of segments for higher level tasks, while the second criterion ascertains that the results of images being segmented accommodate their spatial complexity.

Chapter 6
Synthesis of Intensity Gradient and Texture Information for Efficient Three-Dimensional Segmentation of Medical Volumes

Volumetric segmentation from a medical imaging view-point is defined as the process of segregating an all volumetric pixels (or voxels) in a medical volume into distinct sub-volumes exhibiting (also known as segments or groups or partitions) exhibiting homogeneous or nearly homogeneous attributes. This fundamental data analysis procedure is employed as a pre-processing step for enhancing the performance of numerous higher-level operations such as three-dimensional (3-D) reconstruction, modeling, visualization and navigation, by providing a pertinent representation of the input volume. The increase in number of medical applications has necessitated the development of algorithms that effectively perform segmentation in a timely manner. In this chapter, we propose a novel multi-feature 3-D volumetric segmentation algorithm that efficiently synthesizes intensity (or radiodensity in case of CT), gradient and textural cues, inherent in medical imagery. An overview of our four module 3-D segmentation scheme comprising is shown in Fig. 6.1.

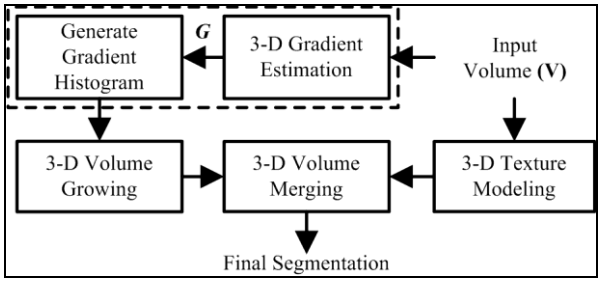

Fig. 6.1: Overview of the proposed 3-D segmentation algorithm.

The first module performs a 3-D gradient estimation technique on the input volume. The second module is responsible for the formation of an initial set of partitions via a 3-D volume growing scheme. The third module is a texture characterization module, which discriminates different textures through an entropy-based descriptor. Finally, the algorithm concludes its operation in the fourth module which performs a 3-D volume merging procedure to arrive at the final segmentation result. The following sub-sections describe each of the aforementioned modules in detail.

6.1. 3-D Vector Field Gradient Estimation

In order to ensure that the proposed segmentation algorithm yields results that are consistent with the intensity variations across the volume, we formulate a 3-D gradient detection scheme based on the work in [Lee and Cok, 1991] which guides the directions in which the volume growing process transpires. In

what follows, let $V_{d\to 3}$ be a 3-D vector field with dimensions KxLxM with d attributes where $a_t \in \{1,2,...,d\}$) denotes the attribute index. Moreover, let x, y, z represent voxel co-ordinates, such that $1 \le x \le K$, $1 \le y \le L$, and $1 \le z \le M$. Consequently, a volumetric gradient matrix (J) also known as the 3-D Jacobian, at voxel location (x, y, z) can be defined as:

$$J(x,y,z) = \begin{bmatrix} \frac{\partial V_1(x,y,z)}{\partial x} & \frac{\partial V_1(x,y,z)}{\partial y} & \frac{\partial V_1(x,y,z)}{\partial z} \\ \vdots & \vdots & \vdots \\ \frac{\partial V_d(x,y,z)}{\partial x} & \frac{\partial V_d(x,y,z)}{\partial y} & \frac{\partial V_d(x,y,z)}{\partial z} \end{bmatrix} \tag{6.1}$$

The gradient magnitude is acquired by first constructing an outer product matrix $J^T J$, given by:

$$J^T J(x,y,z) = \begin{bmatrix} a(x,y,z) & d(x,y,z) & e(x,y,z) \\ d(x,y,z) & b(x,y,z) & f(x,y,z) \\ e(x,y,z) & f(x,y,z) & c(x,y,z) \end{bmatrix} \tag{6.2}$$

where,

$$a(x,y,z) = \sum_{a_t=1}^{d} \left(\frac{\partial V_{a_t}(x,y,z)}{\partial x} \right)^2 \tag{6.3}$$

$$b(x,y,z) = \sum_{a_t=1}^{d} \left(\frac{\partial V_{a_t}(x,y,z)}{\partial y} \right)^2 \tag{6.4}$$

$$c(x,y,z) = \sum_{a_t=1}^{d} \left(\frac{\partial V_{a_t}(x,y,z)}{\partial z} \right)^2 \tag{6.5}$$

$$d(x,y,z) = \sum_{a_t=1}^{d} \left(\frac{\partial V_{a_t}(x,y,z)}{\partial x} \right) \left(\frac{\partial V_{a_t}(x,y,z)}{\partial y} \right) \tag{6.6}$$

$$e(x,y,z) = \sum_{a_t=1}^{d} \left(\frac{\partial V_{a_t}(x,y,z)}{\partial x} \right) \left(\frac{\partial V_{a_t}(x,y,z)}{\partial z} \right) \tag{6.7}$$

$$f(x,y,z) = \sum_{a_t=1}^{d} \left(\frac{\partial V_{a_t}(x,y,z)}{\partial y} \right) \left(\frac{\partial V_{a_t}(x,y,z)}{\partial z} \right) \tag{6.8}$$

Following this, the magnitude of the gradient (G) at voxel location (x, y, z) is computed by finding the square root of the largest Eigenvalue of $J^T J(x,y,z)$, given by:

$$G(x,y,z) = \sqrt{\arg\max_i \{\lambda_i [J^T J(x,y,z)]\}} \tag{6.9}$$

where $\lambda_i [J^T J(x,y,z)]$ is the i^{th} Eigenvalue of $J^T J(x,y,z)$. In Eq. (9) it can be seen that $J^T J(x,y,z)$ cumulatively has three Eigenvalues ($\lambda_i, i = 1,2,3$) calculated specifically at the voxel location (x, y, z).

Also, since $J^T J(x,y,z)$ is a real, symmetric and positive semi-definite 3x3 matrix for every voxel location, the close form solutions of its three Eigenvalues can be given as [Smith, 1961][Zwillinger, 1981]:

$$\lambda_1(x,y,z) = m(x,y,z) + 2\sqrt{p(x,y,z)}\cos\phi(x,y,z) \tag{6.10}$$

$$\lambda_2(x,y,z) = m(x,y,z) - \sqrt{p(x,y,z)}(\cos\phi(x,y,z) + \sqrt{3}\sin\phi(x,y,z)) \tag{6.11}$$

$$\lambda_3(x,y,z) = m(x,y,z) - \sqrt{p(x,y,z)}(\cos\phi(x,y,z) - \sqrt{3}\sin\phi(x,y,z)) \tag{6.12}$$

where,

$$m(x,y,z) = (1/3)[a(x,y,z) + b(x,y,z) + c(x,y,z)] \tag{6.13}$$

$$p(x,y,z) = (1/6)\sum_{i=1}^{3}\sum_{j=1}^{3}\{[J^T J(x,y,z) - m(x,y,z)I_3]^2\}_{ij} \tag{6.14}$$

$$\phi(x,y,z) = (1/3)tan^{-1}[\sqrt{p(x,y,z)^3 - q(x,y,z)^2}\,/\,q(x,y,z)] \tag{6.15}$$

$$q(x,y,z) = (1/2)det[J^T J(x,y,z) - m(x,y,z)I_3] \tag{6.16}$$

Here I_3 is a 3x3 identity matrix, $p(x,y,z)$ represents the sum of squares of the elements of $J^T J(x,y,z) - m(x,y,z)I_3$ weighted by a factor of 1/6, , and det denotes the determinant operator. It can be shown that Eq. (6.10) is indeed the closed form solution of the largest Eigenvalue of $J^T J(x,y,z)$. In Fig. 6.2(a) shown are two slices from a sample MRI dataset of dimensions 256 X 256 X 63. Fig. 6.2(b) reveals corresponding edge information derived by subjecting the entire volume to the above described vector field gradient estimation algorithm. Fig. 6.2(c) portrays edge information rendered as 3-D iso-surfaces (left) in three gradient ranges (see legend) along the x, y, z axis accompanied by a cross-sectional view (right) of 10 slices (#30 through #40) in the $x-y$ plane. Having established the gradient magnitude (G) at all voxel locations in V , a 1-D histogram of these gradient values is generated which is employed in subsequent processing (see Fig. 6.1).

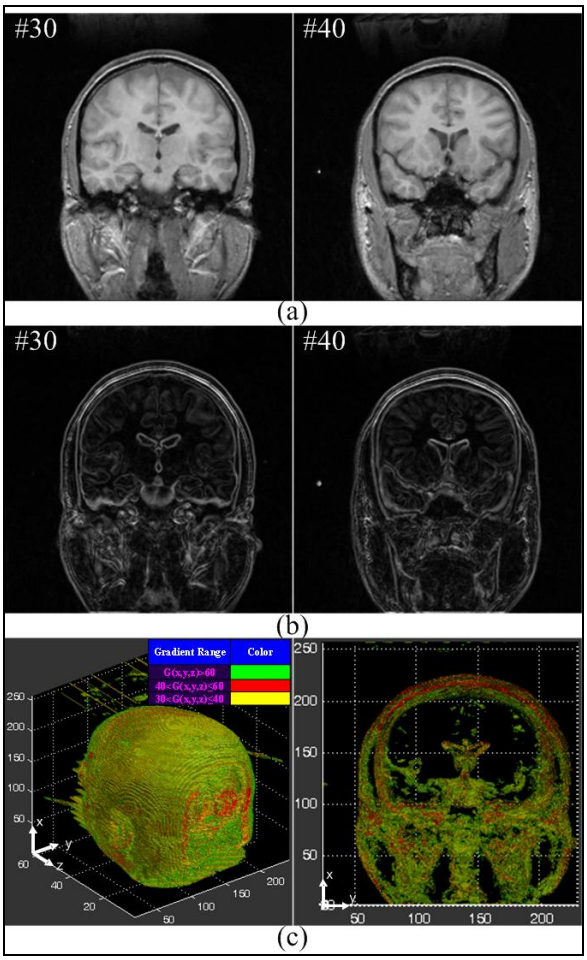

Fig. 6.2: 3-D Gradient estimation: (a) Original MRI slices (#30 and #40), (b) corresponding edge maps, and (c) 3-D view (left) accompanied by a cross-sectional view (right).

6.2. Initial Clustering and 3-D Volume Growing

The proposed algorithm delineates various sub-volumes within V, in two stages namely: initial clustering and volume growing, as shown in Fig. 6.3. The initial clustering step commences the partitioning process by identifying voxel-congregations in low gradient "hollow areas" possessing gradually varying intensity. Such a collection of homogeneous voxel-groupings are identified by choosing an initialization threshold (g_{init}) in a low gradient range, accomplished by computing g_{init} as a small percentage (α) of the maximum value ($G_{\max}$) in G (represented as $g_{init} = \alpha G_{\max}$). Next, voxels in G that satisfy $G \le g_{init}$ are made to undergo connected component analysis [Haralick and Shapiro, 1992], to yield a Parent Seeds (PSs) map (see Fig. 6.4(a)) comprised of several uniquely labeled voxel-clusters that serve as starting points for growth. It is imperative to note that in Fig. 6.4, seeds are randomly color coded for distinction. Furthermore, to avoid managing numerous small seeds in homogeneous areas, a Minimum Seed Volume (*MSV*) criterion is utilized to restrict the volumetric extent of PSs, as depicted in Fig. 6.4(b). *MSV* defines

the minimum number of voxels that every seed should mandatorily contain. This measure is set (empirically at 5 voxels) such that it is small enough to ensure that reasonably fine entities in the volume are captured as the algorithm progresses. Additionally, to ascertain proper sub-volume formation, *PSs* aimed at capturing low gradient locations of the volume are restricted to a large constraint of ~50 times *MSV*, while this criterion is gradually reduced for addition of new seeds at subsequent higher gradient densities at which finer details in the volume manifest themselves. Generation of *PSs* concludes the initial clustering stage, which is succeeded by a 3-D iterative volume growing stage.

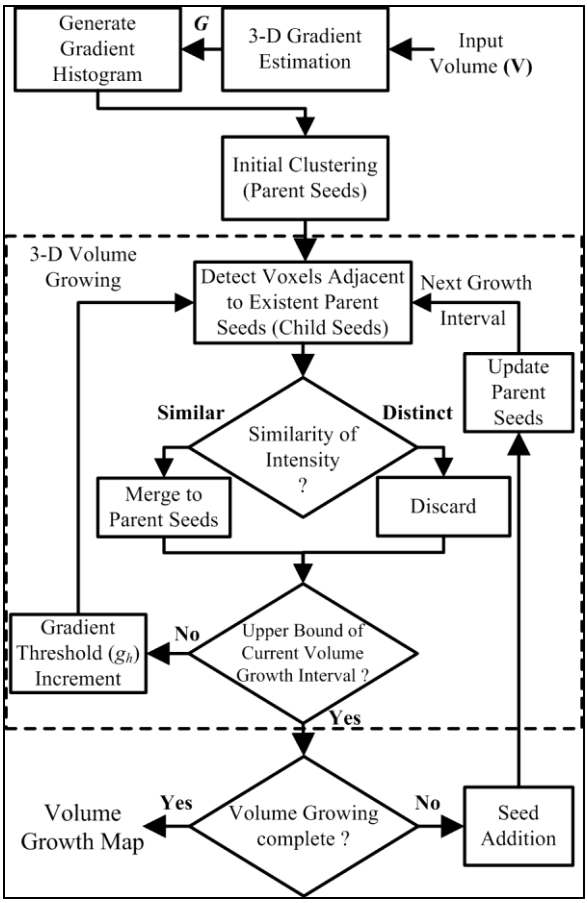

Fig. 6.3: Flowchart of the proposed volume growing method

Volume growing is defined as a technique which starts from a single voxel or a set of individual voxels (termed seeds), and based on a user-defined similarity rule iteratively accumulates voxels around each of them. The growth of a volumetric entity ceases when no more voxels satisfying the aforestated similarity principle are found, and helps form compact 3-D partitions. Most conventional volume growing approaches suffer from some or all of the following short comings: (a) they require manual selection of seeds, (b) they perform voxel-by-voxel agglomerations, (c) they have a sequential architecture (handle only one component of the volume at a time), and (d) they tend to have processing order dependency.

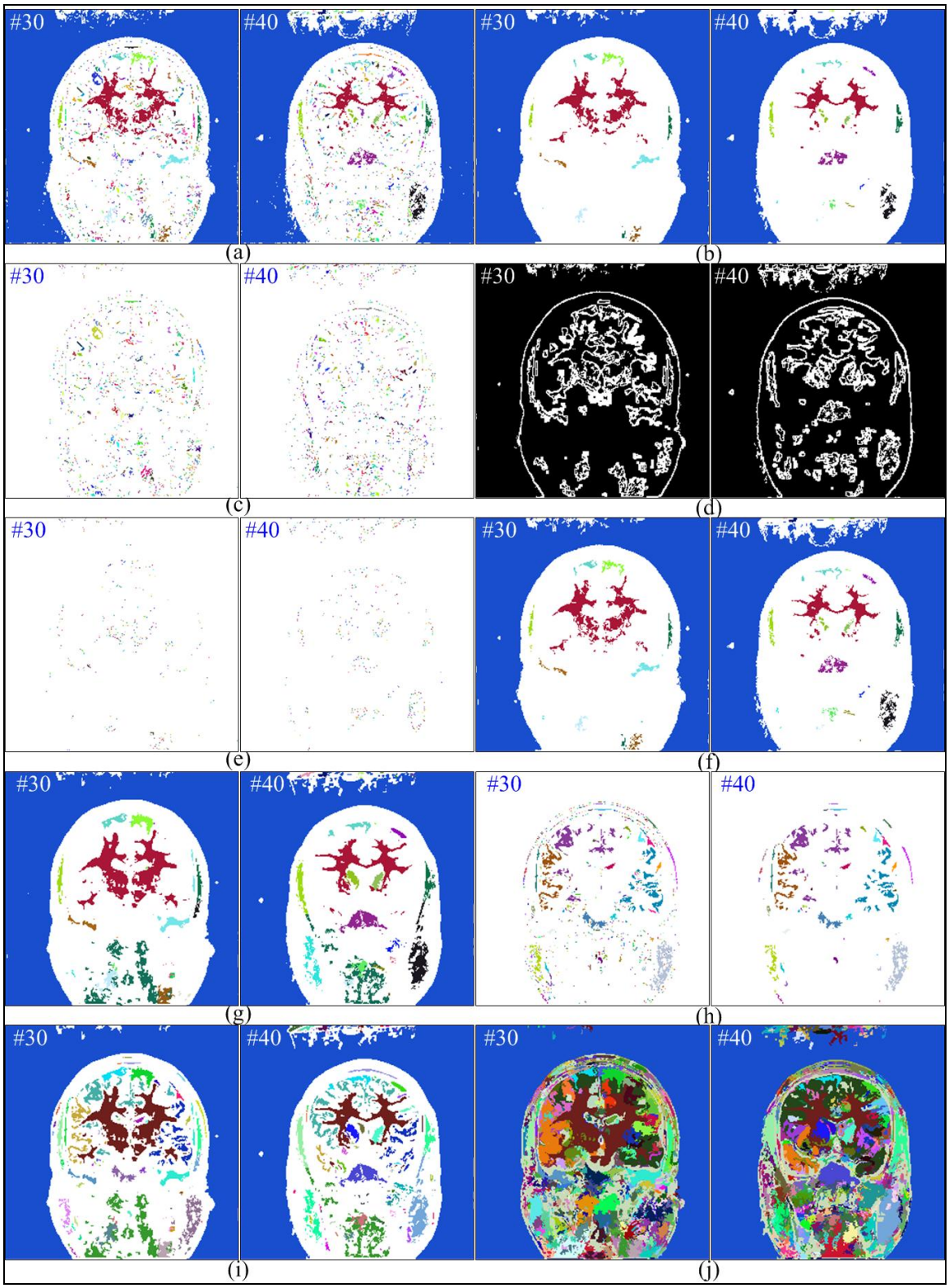

Fig. 6.4: Initial clustering: (a) Initial parent seeds (*PSs*) map using g_{init}=2.5 %{ G_{max}}, and (b) Spatially constrained *PSs* map using 50**MSV*=250 voxels. Volume growing: (c) Child seeds (*CSs*) map, (d) Parent seed surfaces ($PS_{surfaces}$), (e) Adjacent *CSs*, (f) Partition map obtained after parent-child similarities have been processed, (g) partition map obtained at the end of a few iterations of the growth process, (h) New

seeds non-adjacent to existent *PSs*, (i) *PSs* for next stage of volume growing, and (j) Final volume growth map.

Conversely, we propose a volume growing approach that simultaneously lays emphasis on the homogeneous and heterogeneous characteristics of elements in V, using intensity and gradient attributes. Our growth-routine commences at voxels with small gradient magnitudes (intensity homogeneous sections) and terminates at locations with large gradient magnitudes (abrupt intensity transitions) with no dependency on the order in which they are handled. Moreover, at various stages of processing our growing scheme is integrated with an active seed addition mechanism that automatically identifies new seeds in un-segmented portions of V, consequently facilitating parallel growth of multiple adjacent and/or non-adjacent sub-volumes.

The volume growing process proceeds to grow PSs by first identifying the next higher gradient magnitude $g_h > g_{init}$, in the histogram of G. Following this, voxels whose gradient values are less than or equal to g_h and not tagged as PSs, are subjected to connected component labeling to yield a map comprising of new seeds (Fig. 6.4(c)), called Child Seeds (CSs). At this stage, only the CSs adjacent to existent PSs are operated on, identified by distinguishing the ones that share voxels with the peripheral 3-D surfaces of PSs. Parent Seed surfaces ($PS_{surfaces}$) are morphologically derived (Fig. 6.4(d)) as a binary volume by subtracting PSs from their dilated equivalents, represented as:

$$PS(x,y,z)_{surfaces} = \begin{cases} 1 \forall\ (x,y,z)\ where\ \{(PSs \oplus \psi) - PSs\} > 0 \\ 0 \forall\ (x,y,z)\ where\ \{(PSs \oplus \psi) - PSs\} = 0 \end{cases} \tag{6.17}$$

In Eq. (6.17), $PSs \oplus \psi$ represents the 3-D dilation of the PSs map with a uniform (all 1's) 3x3x3 cubic structuring element ψ. Following this, the labels of CSs bearing adjacency with PSs are obtained (Fig. 6.4(e)) by multiplying the $PS_{surfaces}$ and CSs maps on a per-voxel basis. Furthermore, to avoid computational burden, adjacent CSs whose sizes are smaller than the *MSV* criterion are directly assigned their associated parent labels, while the larger ones are spectrally compared to their parents. Parent-child similarity is assessed by computing the Euclidean distance ($\Delta E_{PSs-CSs}$), between their mean intensity values (m_{PSs}) and (m_{CSs}), denoted as:

$$\Delta E_{PSs-CSs} = \sqrt{(m_{PSs} - m_{CSs})^2} \tag{6.18}$$

To this effect, the adjacent CSs (larger than the *MSV* measure) whose mean intensity values are found to be comparable to their parents are merged with them, while the rest are discarded, as shown in Fig. 6.3. Specifically, the aforestated parent-child similarity is inferred by assessing whether the obtained $\Delta E_{PSs-CSs}$ values are less than a maximum allowed distance ΔE_{max} (set equal to 1). This culminates in a single

iteration of volume growing at a gradient threshold g_h, resulting in the gradual growth of PSs (see Fig. 6.4(f) relative to Fig. 6.4(b)). Additionally, as the aforementioned growing procedure advances iteratively into higher gradient densities effecting parallel growth of all PSs, volumetric compartments that are non-adjacent to them remain un-segmented in the partition map, as portrayed in Fig. 6.4(g). Therefore, at specific processing intervals our growth mechanism is interleaved with an active seed addition step to facilitate addition of new seeds in non-adjacent expanses to the existent PSs. These intervals, termed as volume growth intervals, are derived by dividing the entire gradient histogram into M components demarcated using $M-1$ gradient values, such that each component spans approximately equal area (or number of voxels). Fig. 6.5(a) shows the 1-D gradient histogram of an MRI volume previously displayed in Fig. 6.2(a), fragmented into ten volume growth intervals (distinguished using red lines) each spanning equal (~10%) area. Additionally, Fig. 6.5(b) portrays a zoomed portion of the histogram (boxed in green in Fig. 6.5(a)) for a clear view of the various growth intervals. The aforestated choice of the number of intervals was found to optimally control (i.e. not being too many or too few) the number of seeds being added to the partition map, whilst accounting for the exponential nature of the gradient histogram. Thus, on conclusion of the growth process in an arbitrary interval (when g_h is equal to its upper bound), voxel-clusters in expanses that are not adjacent to existing seeds and relatively larger than the MSV criterion (Fig. 6.4(h)) are added to the region map. The resulting agglomeration of seeds becomes the PSs map for the next interval of volume growing, as seen in Figs. 6.3 and 6.4(i). This interlaced mechanism involving volume growing and active seed addition culminates when the gradient threshold equivalent to the highest magnitude $G_{\max}$ is dealt with, to arrive at the final volume growth map (Fig. 6.4(j)).

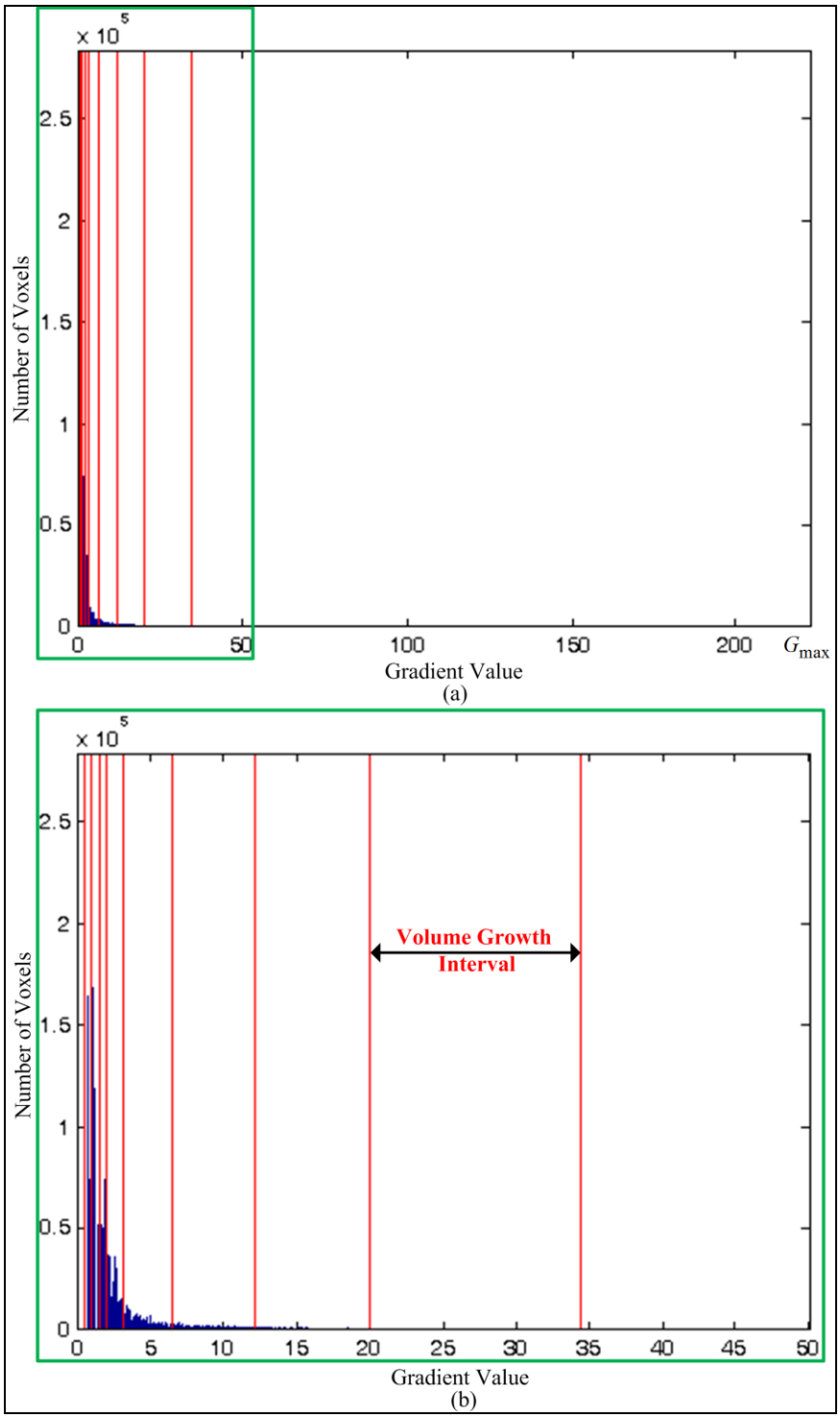

Fig. 6.5: (a) Gradient histogram of a sample MRI dataset (shown in Fig. 6.2(a)) divided into ten volume growth intervals each spanning equal area, and (b) portion of the gradient histogram boxed in green (displayed for viewing ease).

6.3. 3-D Entropy-based Texture Modeling

Contrasting intensity fluctuations ranging from multiple shades of intensity to repeating patterns of local intensity variations, pertinent to organs, soft-tissues, bones, air, and liquid substances are abundantly present in medical volumes. These fluctuations pose significant problems in achieving the segmentation objective, often resulting in over-segmentation and misinterpretation or perceptual ambiguity of boundaries/ surfaces surrounding them. Such volumetric zones of intensity variations are referred to as textures and may contain structured or stochastic patterns/distributions. To mitigate this problem, the proposed algorithm was furnished with a texture modeling component (see Fig. 6.1) which characterizes different textures via a 3-D entropy-based descriptor.

For a 3-D input MRI/CT scalar field, utilizing Eq. (3.17) an entropy calculation is done locally in a 3x3x3 neighborhood around every voxel location to arrive at a descriptor that distinguishes various textures with in the input volume. Additionally, since medical image data typically contains high bit-depths (12-16 bits) of information per voxel, the resultant probability distributions of various intensities tend to be quite large and sparsely populated. To this effect, the input volume is quantized to low bit-depths (or few gray levels) before being employed for texture characterization to reduce computational load. The above mentioned quantization step can also be employed in case of 3-D vector fields to avoid the computation of the joint probability distribution of intensities across attributes for entropy calculation, as seen in Eq. (3.18).

6.4. 3-D Volume Merging using Analysis of Variance

The set of sub-volumes formed at the culmination of the growing process is typically an over-segmented result with reference to the composition of the input volume (see Figs. 6.2(a) and 6.4(j)). This is due to the fact that these partitions are assembled across various gradient densities in a process that, although accounts for the intensity similarity of voxels, is oblivious to the abundant presence of texture disparities among them. Hence, the last module in our segmentation framework involves a 3-D volume merging routine that unifies sub-volumes with similar characteristics in a multivariate intensity-texture space, to arrive at the final/refined set of segments. The aforesaid functionality is accomplished through the Multivariate Analysis of Variance (MANOVA) procedure, which helps construct a canonical variate space to emphasize between-group differences, and ultimately scrutinize the possibility of relating various sub-volume pairs to a single factor that enables them to be combined together.

On the commencement of the merging process the sub-volume-pair with the minimum distance, signifying the maximum similarity is identified. However, instead of fusing only a single sub-volume-pair per merging iteration, we adopt a “greedy” algorithm for computational efficiency (as described previously), where the distance between the two most similar partitions is gradually increased until a

larger set of "highly similar" sub-volume pairs are obtained, before merging is performed. In this subset of similar sub-volume pairs, partitions are fused with each other from the most similar pair to the least similar one, which culminates a single volume merging-iteration. Next, the distances are recomputed for all possible pairs in the new/refined set of sub-volumes. This process is repeated hierarchically until, in the specified order of priority, either the distance value exceeds a pre-defined threshold (T =1) or the desired number of segments (S set based on complexity of the volume) are obtained, yielding the final 3-D segmentation outcome. While the first criterion ensures automatic acquisition of a practicable number of segments for higher level tasks, the second rule ascertains that the proposed algorithm accounts for the complexity of the input data.

6.5. Semi-Automatic 3-D Segmentation using the Proposed Framework

Although MRI/CT and other medical imaging systems continuously capture tremendous volume of data, extraction of meaningful diagnostic information from the acquired datasets rests squarely on the shoulders of the end users. For example, radiologists are often interested in tools that facilitate quick and effective 3-D visualization of a desired volumetric entity such as an abdominal organ or vertebral structure, to aid sound medical reasoning and judgment. Consequently, in this section, we discuss a semi-automatic segmentation algorithm derived as a subset of the segmentation framework proposed in Section II that enables a user to interactively segment a desired Sub-Volume of Interest (SVOI), as opposed to the entire volume. An overview of the algorithm consisting of three modules is shown in Fig. 6.6. The first module is responsible for the SVOI selection procedure. The second module performs 3-D gradient detection across the volume. Finally, the third module performs volume growing to arrive at the final segmentation result of the specified SVOI. The following sub-sections describe the functionalities of each of the three modules.

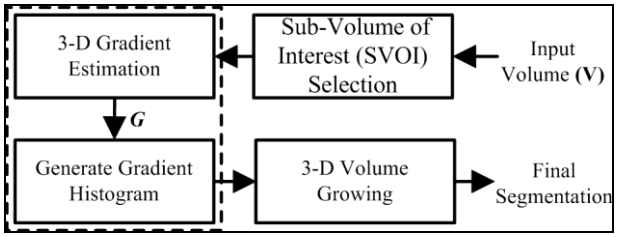

Fig. 6.6: Overview of the proposed semi-automatic segmentation algorithm.

6.5.1 Sub-Volume of Interest (SVOI) Selection

The algorithm commences interactively, where-in a user indicates the SVOI by pair-wise selecting the extremities of a few 'line markings' within its cross-section, in an arbitrary slice (Fig. 6.7(a)) of the input CT volume. Also, shown are user-selected start (x_1, y_1, z_1) and end (x_2, y_2, z_1) voxels of a single line

marking (l_1) in slice z_1, inside the left lung. Given two such extremities in a plane/slice z_1, the general equation of the line traversing them can be obtained as:

$$y - y_1 = m(x - x_1) \Rightarrow mx - y + (y_1 - mx_1) = 0 \tag{6.19}$$

where the slope of the line $m = (y_2 - y_1)/(x_2 - x_1)$. Consequently, a voxel (x_i, y_i, z_1) in the plane z_1 lies on the line represented by Eq. (6.19) if and only if:

$$mx_i - y_i + (y_1 - mx_1) = 0 \tag{6.20}$$

The set of all voxels satisfying Eq. (6.20) (shown as green dotted lines in Fig. 6.7(b)) and constrained within (x_1, y_1, z_1) and (x_2, y_2, z_1) define the line marking l_1, as seen in Fig. 6.7(c). Subject to the complexity of the SVOI, single or multiple such line markings are utilized to effectively initialize the volume growing procedure.

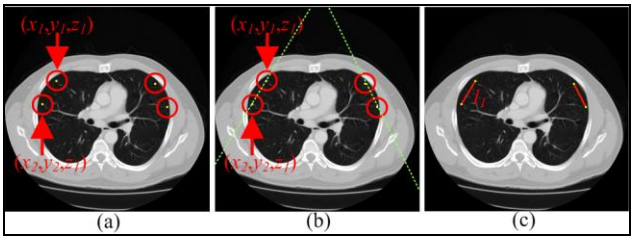

Figure 6.7: Sub-Volume of Interest selection

6.5.2 Gradient Estimation and Volume Growing

Once the SVOI is selected by a user, a 3-D gradient volume (G) is computed from the input data using the scheme described in Section 6.1. This is followed by the generation of its histogram, which is employed in the subsequent initial clustering and volume growing steps, utilized to segment the SVOI from the input volume. The initial clustering step instigates formation of the SVOI by identifying a group of voxels possessing similar or gradually varying intensity in the vicinity of the user-specified 'line markings' (internal to the SVOI). To ensure that such a set of homogeneous voxels manifesting small gradient magnitudes among them are isolated, an initialization threshold (g_{init}) equivalent to the most commonly prevalent gradient value (given by the mode) in all line markings, is chosen. In general, if $l = \{l_1, l_2, ..., l_N\}$ denotes a set of N user indicated line markings each of length $\{n_1, n_2, ..., n_N\}$ number of voxels respectively, and $G_l = \{G_1, G_2, ..., G_N\}$ represents the set of gradient magnitudes associated with l, where G_i signifies the gradient magnitudes of voxels that define l_i, then g_{init} is selected such that:

$$g_{init} = \begin{cases} mode(G_l) & if \quad mode(G_l) \ is \ unimodd \\ \gamma : \gamma \in mode(G_l) & if \quad mode(G_l) \ is \ multimodal \end{cases} \tag{6.21}$$

When the mode of G_l is not unique (multimodal), a random choice γ from the acquired multiple mode values is used for g_{init}, as represented in Eq. (6.21). From this point forward, the 3-D volume growing

routine described in Section 6.2 is iteratively carried out over all gradient magnitudes (higher than g_{init}) which culminates when the gradient threshold equivalent to the maximum magnitude in G is processed, to arrive at the final 3-D segmentation result of the SVOI. It is important to observe that the active seed addition component of the volume growing process, as well as the texture characterization and volume merging modules are not mandated, given that the semi-automatic algorithm is targeted only at a single SVOI uniquely seeded and labeled via the user defined line markings.

From a computational efficiency stand point both the automatic and semi-automatic 3-D segmentation algorithms have been implemented entirely in a vectorized fashion. To this effect, for an arbitrary medical volume of dimensions ~512x512x150 the run time of the algorithm is about ~2-4 minutes in a MATLAB environment in a standard desktop setting. A GPU implementation of the aforementioned segmentation framework (as eventually done with most state-of-the-art algorithms) would significantly reduce the computational load (to the order of a few seconds), making it suitable for real time use.

Chapter 7

Experimental Results and Discussions

This chapter demonstrates the experimental results achieved over the course of this doctoral study. Section 7.1 and Section 7.2 provide the color image segmentation results from the GSEG and MAPGSEG algorithms (first and second research objectives). Section 7.3 depicts the results of our proposed algorithm for segmenting multi/hyperspectral data (third research objective). Finally, Section 7.4 discusses the results achieved from our 3-D segmentation algorithm demonstrated on MRI/CT volumes.

7.1. GSEG Algorithm Results

Until recently, different segmentation algorithms demonstrated their effectiveness by displaying results obtained on a limited set of images. On the contrary, to objectively measure the quality of our segmentation results, we have chosen a recently proposed evaluation metric called the Normalized Probabilistic Rand (NPR) index. The NPR evaluation method reported in [Unnikrishnan et al., 2007] is summarized in Appendix B. The essence and effectiveness of the NPR lies in the fact that it compares results obtained from a tested algorithm to a set of multiple manually segmented ones for a given image taking into consideration the variability across the manual segmentations. Furthermore, this set is imperative for segmentation evaluation, as there is not a unique correct outcome for an image; consequently, the set of multiple perspectives of correct segmentations becomes the corresponding ground truths for that image. The NPR technique is very generic and is not application specific like some of the earlier ones, modeled to account for variability in human perception. Due to the fact the calculated NPR index varies depending on the image's segmentation ground-truth utilized in the evaluation process, they are required to display the following characteristics: 1) they cannot be chosen selectively to favor a given algorithm, 2) they display various scenarios with multiple levels of complexity, 3) they contain more than one individual perspective, and 4) they can be accessed by anyone for performing the same test on diverse algorithms.

Such a set has been made publicly available by the University of California at Berkeley. This database provides 1633 manual segmentations for 300 images created by 30 human subjects [Martin et al. 2001]. Moreover, for our evaluation, state-of-the-art algorithms were chosen to furnish a fair indication of the quality of segmentation results achieved. These segmentation techniques are fusion of color and edge information for improved segmentation and edge linking (GRF) [Saber et al., 1997], the unsupervised segmentation of color-texture regions in images and video (JSEG) [Deng et al., 2001], as well as the Efficient Graph-based Segmentation (EGS) method introduced by [Felzenszwalb et al., 2004]. To prevent

any discrepancies, all the available images were segmented using the above mentioned segmentation algorithms on the same machine. The GRF and JSEG algorithms were run from the executable file provided by the Rochester Institute of Technology and the University of California at Santa Barbara, respectively, while the EGS algorithm was run from an open source C++ implementation provided by Felzenszwalb et al. The proposed method was originally implemented using MATLAB version R2007a.

The results of the GSEG algorithm at different stages are presented in Figs. 7.1(a)-(f). The input RGB image and its CIE L*a*b* counterpart are shown in Figs. 7.1(a) and 7.1(b) respectively. The outcome of gradient computation on the color converted input image is shown in Fig. 7.1(c). The seed map at the end of the region growth procedure, obtained utilizing thresholds that are generated adaptively, is displayed in Fig.7.1(d). Observe that this interim result is over-segmented, due to reasons specified in Section 3.7. The texture channel generated using color quantization and local entropy calculation is depicted in Fig. 7.1(e). The segmentation map at the end of the region merging algorithm is shown in Fig. 7.1(f).

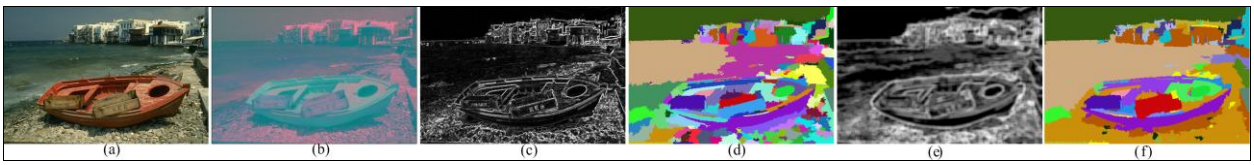

Fig. 7.1: (a) Original RGB image, (b) color space conversion, (c) gradient map, (d) seed map after region growth, (e) texture channel, and (f) final segmentation map after region merging.

Results obtained from the GSEG in comparison to the previously mentioned segmentation methods, are shown in Figs. 7.2 – 7.6. An image with less complex content is shown in Fig. 7.2(a). The image has relatively uniform color information without much variation due to illumination, and the absence of texture in all its regions, renders it to be a simple image to segment. The results shown in Figs. 7.2(b) - 7.2(e), obtained from the GRF, JSEG, EGS and GSEG for this image appear to be similar. Conversely, it can be seen that in case of the GRF, JSEG and the EGS segmentation results, the foliage is partitioned as two different regions in spite of their color of being similar. This is overcome by our new multivariate merging procedure that merges these regions which are almost separate in the image. Clear advantages can be seen on the level of detail achieved in our segmentation results in comparison to the GRF and JSEG, for images with higher complexity.

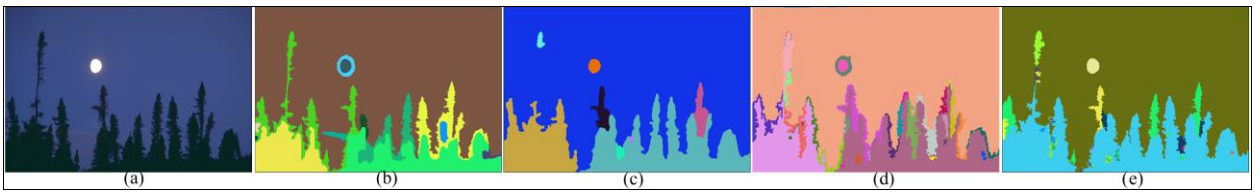

Fig. 7.2: Moon light Results: (a) Original, (b) GRF, (c) JSEG, (d) EGS and (e) GSEG.

The ‘Face’ image in Fig. 7.3(a) represents a moderately complex image with dissimilar texture content associated with the skin, hat and robe of the person. Observe that in Figs. 7.3(b), 7.3(c), and 7.3(d) the GRF, JSEG, and EGS algorithms over segment this image due to the texture and illumination disparity seen in various regions. However, our algorithm employs the CIE L*a*b* color space where the L* channel contains the luminance information in the image, overcomes the illumination problem. Furthermore, our texture model has been effective in handling different textures as seen in Fig. 7.3(e).

Segmenting textured regions becomes a harder challenge when regions with diverse textures are extremely similar in color. In such scenarios, a good texture descriptor is indispensible. Fig. 7.4(a) represents an image of a Cheetah which has a skin tone that almost matches its background making it extremely difficult to segment it based on just color information. The GRF, JSEG, and EGS results shown in Figs. 7.4(b), 7.4(c), 7.4(d) illustrates the effect of an indistinct texture descriptor for segmentation. The GSEG algorithm in comparison (Fig. 7.4(e)) has been able to achieve a good segmentation illustrating the robustness of our texture model in handling highly complex images.

Fig. 7.3: Face Results: (a) Original, (b) GRF, (c) JSEG, (d) EGS, and (e) GSEG.

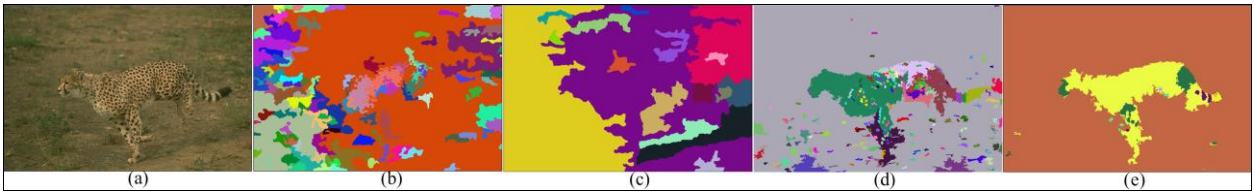

Fig. 7.4: Cheetah Results: (a) Original, (b) GRF, (c) JSEG, (d) EGS, and (e) GSEG.

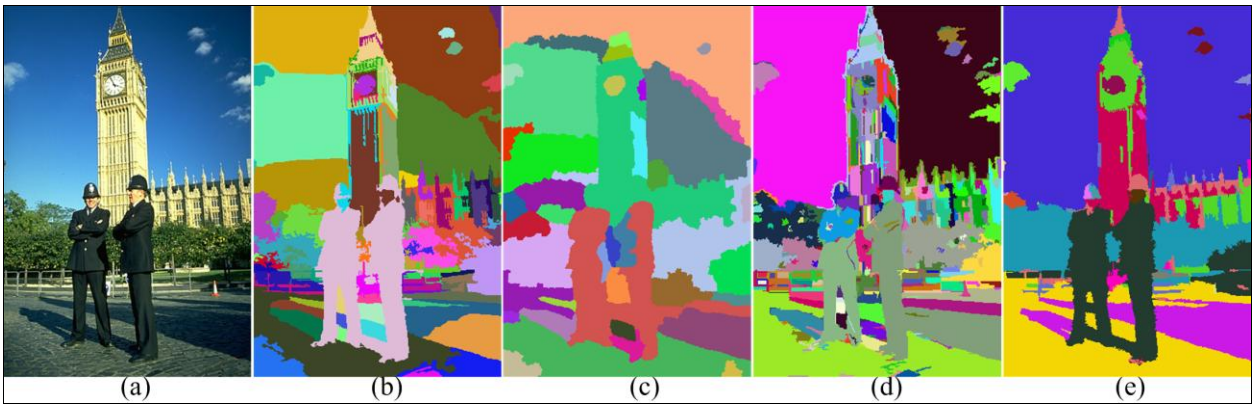

Fig. 7.5: London Results: (a) Original, (b) GRF, (c) JSEG, (d) EGS, and (d) GSEG.

Results obtained for the following set of images demonstrate the advantages of our initial clustering technique to overcome problems of varying illumination, as discussed previously. These images represent instances of largely varying illumination conditions in comparison to previous examples, as well as occlusions by foreground objects. Segmentation results of the 'London' image are depicted in Fig. 7.5. Observe that although the GRF algorithm produces a clear segmentation where the boundaries of the segments match the boundaries of the object as seen in Fig. 7.5(b), it suffers from over segmentation in the sky region. This is due to the use of predetermined number of clusters for segmentation. This phenomenon of over segmentation can also be examined with the JSEG algorithm (see Fig. 7.5(c)) owing to the fact that the initial clusters are based on the quantization of colors that causes varying shades of colors to produce a number of clusters. In this image the sky has been over segmented, because the change of light provides different shades of blue. This problem of varying illumination has yet again been well tackled by the GSEG algorithm which employs a region growth procedure that does not exclusively depend on the initial assignment of clusters, as seen in Fig. 7.5(e). This aspect of the GSEG algorithm has also enabled it to segment even fine details such as text with great efficiency unlike the GRF and JSEG, as illustrated in results in Fig.7.6.

Observe that the word 'Castrol' as seen in Fig. 7.6(a) is segmented out at multiple locations with near perfection by the GSEG algorithm as seen in Fig. 7.6(e). Yet again, the GRF, JSEG, and EGS cause over segmentation in regions representing the motorway due to varying illumination, as seen in Figs. 7.6(b), 7.6(c), and 7.6(d). In addition, the efficiency of the GSEG algorithm in handling the background occlusion problem discussed earlier is emphasized in the 'London' and 'Cars' results. In Fig. 7.5(e) the sky region in spite of being completely occluded by the clock tower has been segmented out as a single region and in Fig. 7.6(e) the motorway has been segmented as one region even though the cars occlude it almost entirely, unlike the results of the GRF, JSEG and EGS in Figs. 7.5(b), 7.5(c), 7.5(d) and Figs. 7.6(b), 7.6(c), 7.6(d). In these results, multiple regions are assigned to the same class, yielding a reduction

in the total number of classes without losing information obtained from multiple similar regions that are not adjacent to each other. This has been possible only due to our unique region merging procedure.

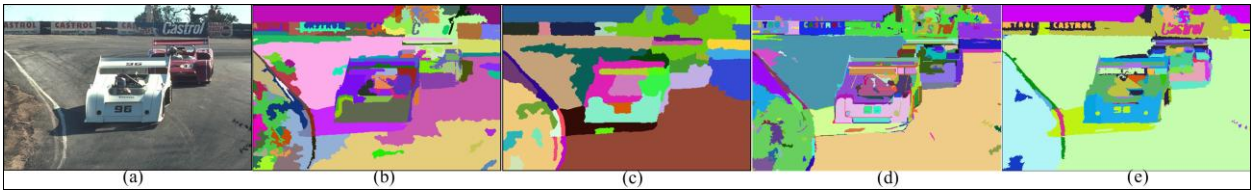

Fig. 7.6: Cars Results: (a) Original, (b) GRF, (c) JSEG, (d) EGS, and (e) GSEG.

The normalization factor was computed by evaluating the Probabilistic Rand (PR) for all available manual segmentations (see Appendix B), and the expected index ($E[PR]$) obtained was 0.6064. A comparison of our evaluation, for the segmentation results obtained from the four methods, is displayed in Table 7.1. This table shows that our algorithm has the highest average NPR score, and the variance of the NPR scores has the narrowest spread, as well as the highest number of images in the range $0.7<NPR<1$ associated with very good segmentation results, illustrating that our algorithm performs consistently better than the other algorithms qualitatively and quantitatively. In addition, the GSEG has an average computation time comparable to the JSEG, and EGS considering the different environments in which they were developed.

The distribution of the NPR scores for the Berkeley database is displayed in Fig. 7.7. This figure illustrates our algorithm peaking at a NPR score of 0.7, ahead of JSEG's 0.6, EGS's 0.6 and GRF's 0.4. When using the NPR scores to create a cumulative percentage of images below a specific NPR score (see Fig. 7.8), the curve representing the GSEG method always stays to the right. This shows that for any NPR score, the GSEG method has a larger number of images segmented with a higher NPR value than the GRF, JSEG and EGS methods.

TABLE 7.1: EVALUATION OF SEGMENTATION RESULTS

	GRF	JSEG	EGS	GSEG
Avg. Time (sec)	240	16.2	7.0	24.1
Avg. NPR	0.358	0.439	0.457	0.495
Std. NPR	0.345	0.318	0.324	0.306
NPR>0.7 (# Images)	39	65	68	91
Environment	C	C	C	MATLAB

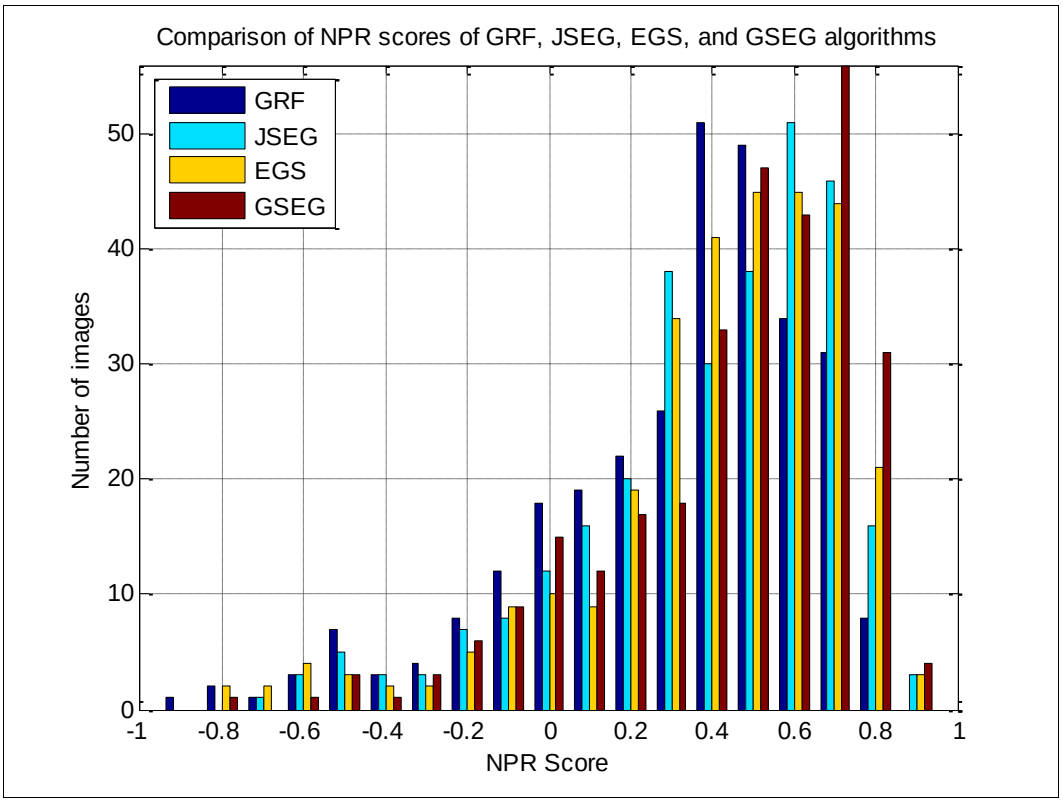

Fig. 7.7: Distributional comparison of NPR scores for 300 images of the Berkeley database.

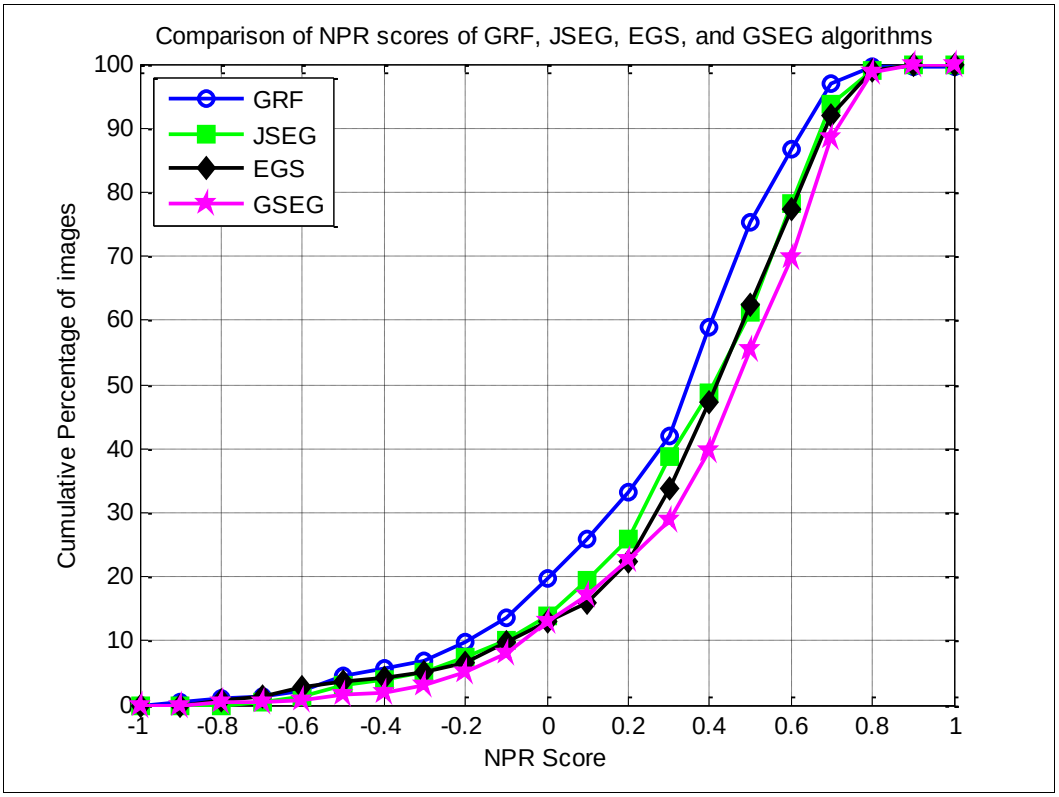

Fig. 7.8: Cumulative Percentage of NPR Scores.

Figs. 7.9-7.12 provide additional results to scrutinize the algorithm's performance on various scenes. Additionally, our algorithm was tested on a database of several hundred images with varying complexity and similar results were observed, indicating the algorithm's robustness to various image scenarios.

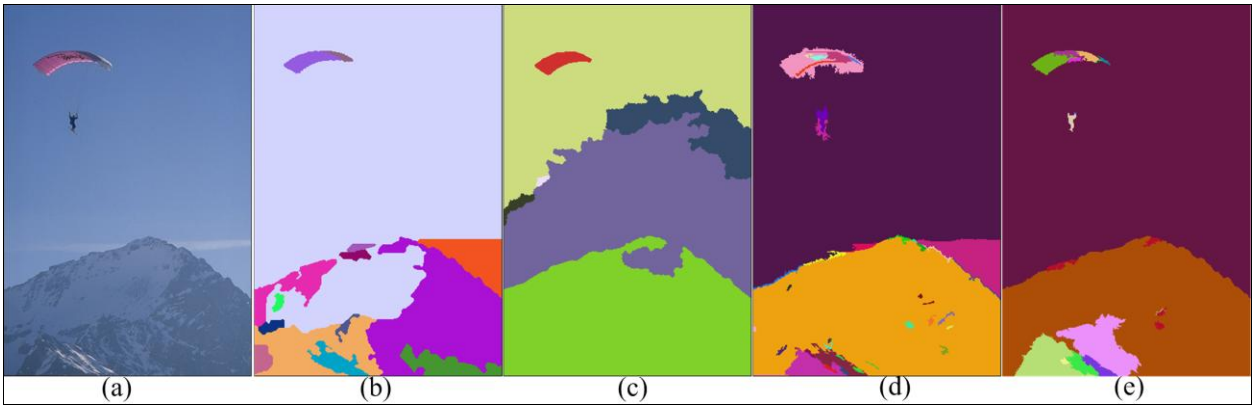

Fig. 7.9: Parachute Results: (a) Original, (b) GRF, (c) JSEG, (d) EGS, and (e) GSEG.

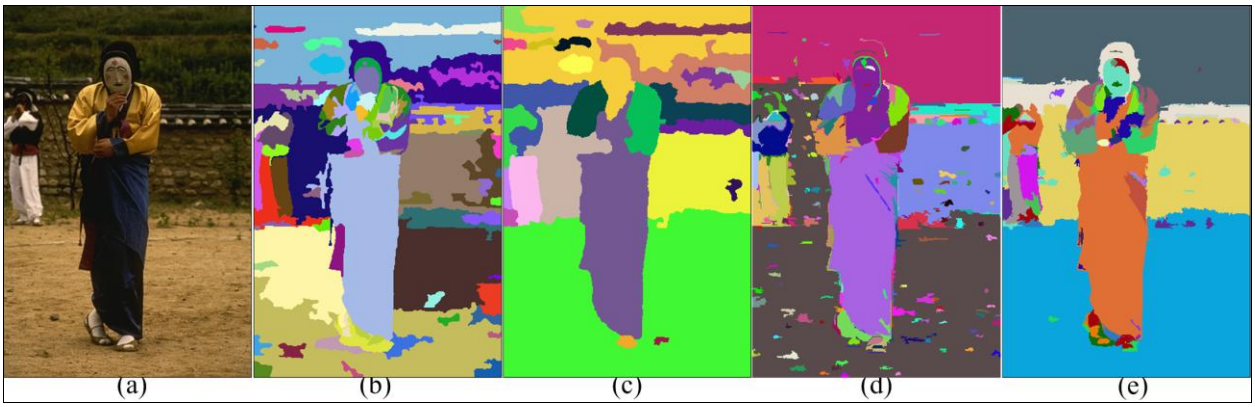

Fig. 7.10: Mask Results: (a) Original, (b) GRF, (c) JSEG, (d) EGS, and (e) GSEG.

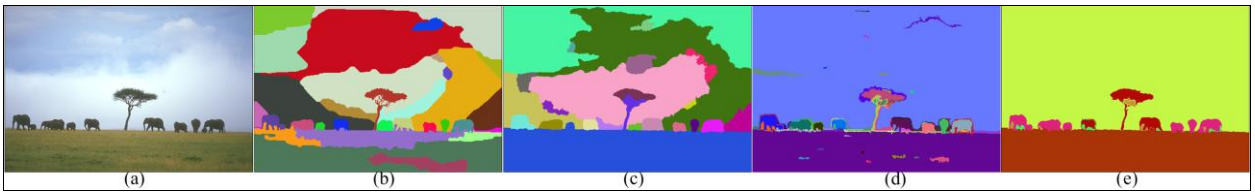

Fig. 7.11: Safari Results: (a) Original, (b) GRF, (c) JSEG, (d) EGS, and (e) GSEG.

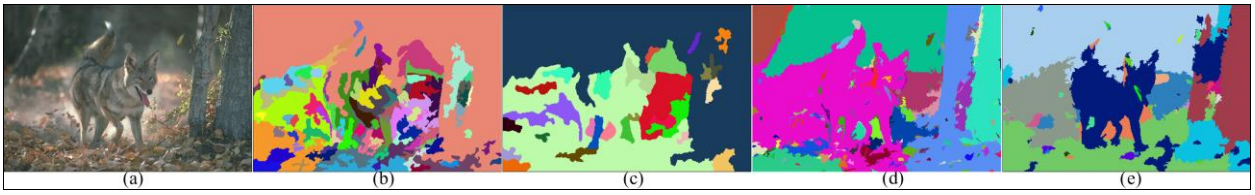

Fig. 7.12: Fox Results: (a) Original, (b) GRF, (c) JSEG, (d) EGS, and (e) GSEG.

7.2. MAPGSEG Algorithm Results

Similar to the GSEG algorithm, the results of the MAPGSEG algorithm were benchmarked qualitatively and quantitatively using the Normalized Probabilistic Rand index (NPR) [Unnikrishnan et al., 2007] on the same test bed of 1633 ground truth segmentations made publically available by the University of California at Berkeley [Martin et al. 2001], against the same set (GRF, JSEG and EGS) of popular segmentation algorithms. The proposed method was originally implemented entirely in MATLAB R2008a.

The results of the MAPGSEG algorithm for a two level decomposition, at different stages of processing, are presented in Figs. 7.13(a)-7.13(f). The approximations of the input RGB image at different dyadic scales and their corresponding color converted counterparts are shown in Figs. 7.13(a) and 7.13(b) respectively. Fig. 7.13(c) displays the outcomes of the vector color gradient computation, at various resolutions. The seed maps at the end of the progressive region growth and distributed dynamic seed addition procedure are displayed in Fig. 7.13(d). Observe that the obtained region growth maps are over-segmented due to reasons specified in Section 4.7, and the extent of over-segmentation decreases with increased scale, such that at the highest resolution (*L=0*) the region growth map is a close representation of the eventual segmentation. This behavior signifies that a large portion of region processing and partitioning occurs at relatively coarse resolutions, consequently resulting in low computational expenses. In Fig. 7.13(e), the texture channels generated using local neighborhood based entropy calculations are depicted. Lastly, the interim segmentations at levels *L=2, 1* and final segmentation map at level *L=0*, after the conclusion of the region merging process at each scale, are shown in Fig. 7.13(f).

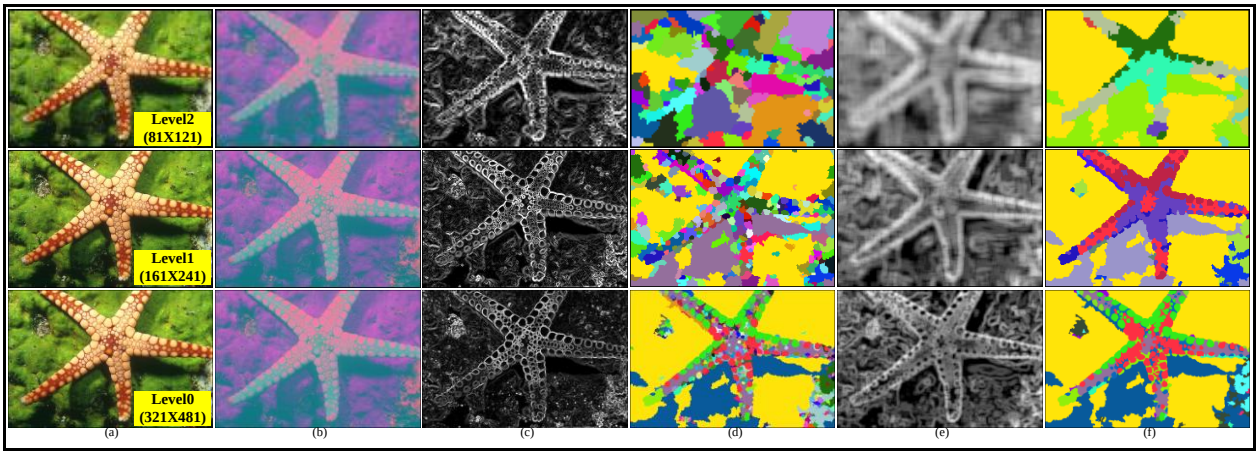

Fig. 7.13: Results of the MAPGSEG algorithm at various stages of processing: (a) original RGB 'Star Fish' image, (b) color converted 'Star Fish' image, (c) color gradient, (d) seeds maps at the end of progressive region growth, (e) entropy based texture maps, (f) interim and final segmentation outputs.

Results acquired from the MAPGSEG algorithm in comparison to the ones obtained from the previously mentioned segmentation methods, are shown in Figs. 7.14 - 7.18. Fig. 7.14 shows the results of a 'skydiver' image (see Fig. 7.14(a)), consisting of well distinct foreground and background regions in gradually varying illumination conditions and slight texture disparities, consequently rendering an image with relatively low complexity. In Figs. 7.14(b), 7.14(c), 7.14(d) it can be observed that GRF, JSEG and EGS algorithms over-segment the sky regions predominantly due to illumination variations. However, the MAPGSEG algorithm that performs region processing in the $L^* a^* b^*$ color space, which constitutes the information rich luminance (L^*) component helps overcome the aforementioned illumination problem, to produce a result with a well segmented sky region, as shown in Fig. 7.14(g).

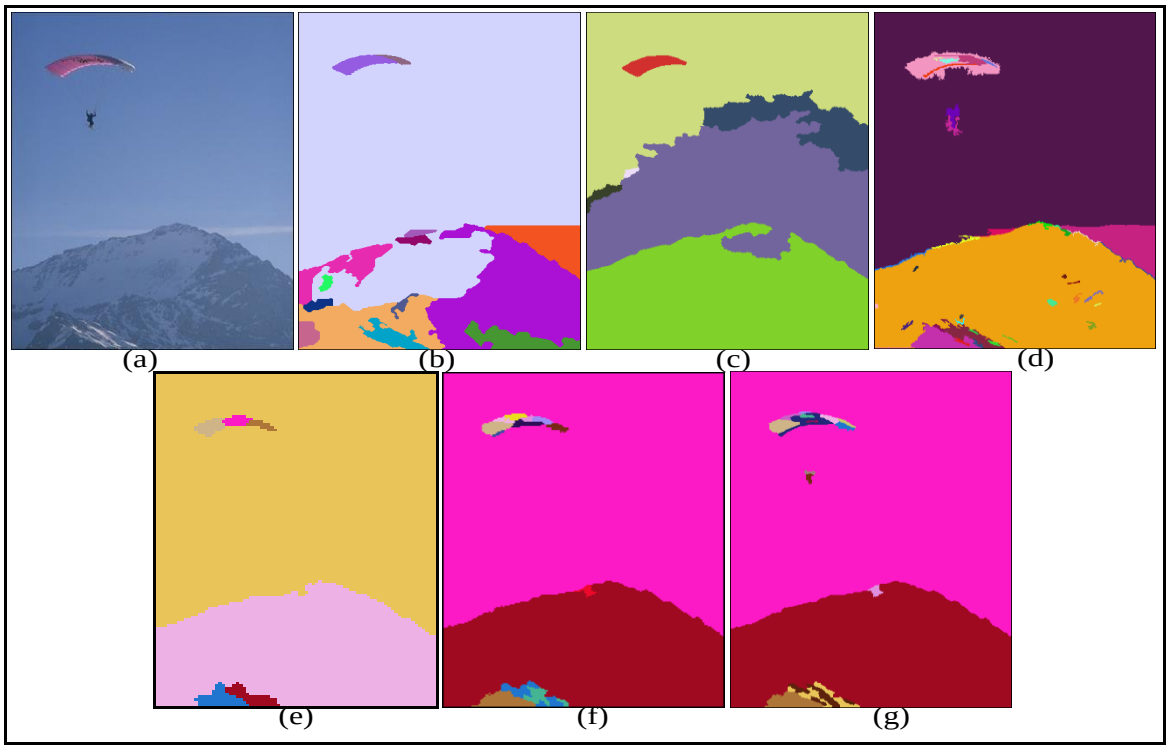

Fig. 7.14: Skydiver results: (a) original RGB image, (b) GRF, (c) JSEG, (d) EGS, and MAPGSEG (e) L=2, (f) L=1, (g) L=0.

In Fig. 7.15 the results of a reasonably complex image of a church (Fig. 7.15(a)), consisting of drastic illumination variations and considerable amounts of structural details, is shown. Observe that in Figs. 7.15(b), 7.15(c), 7.15(d), the GRF, JSEG, and EGS algorithms over segment this image in the sky and dome regions due to illumination disparities. In addition, the presence of numerous structural details causes these algorithms over segment to the façade of the church. However segmentation map displayed in Fig. 7.15(d) demonstrates the efficiency of the proposed algorithm in handling the illumination problem for reasons specified previously. Furthermore the MAPGSEG algorithm equipped with a region

growth strategy that proceeds from color homogeneous regions to regions of color transitions, and a unique region merging procedure that merges similar color-texture regions independent of their spatial locations, results the church being an appropriately segmented.

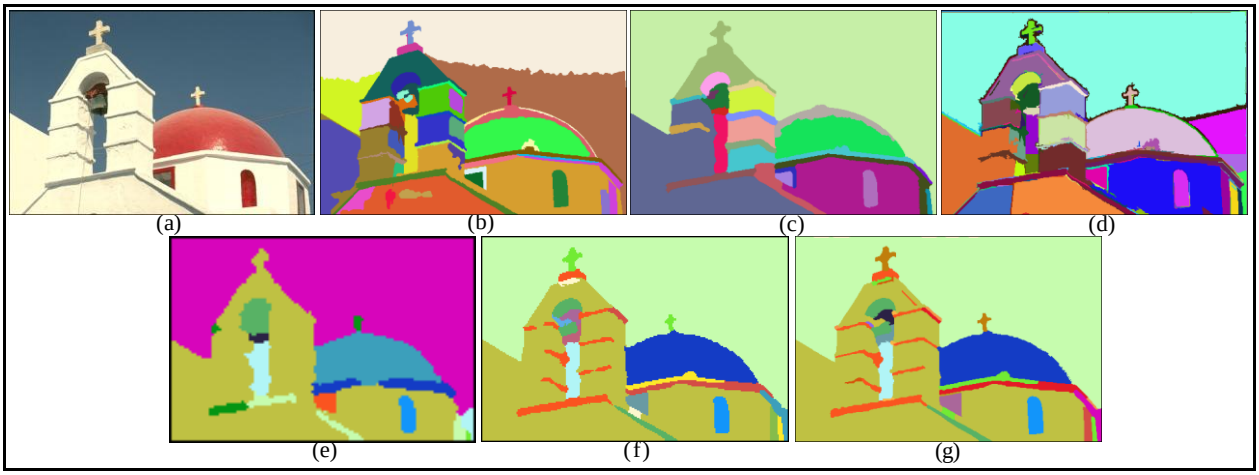

Fig. 7.15: Church results: (a) original RGB image, (b) GRF, (c) JSEG, (d) EGS, and MAPGSEG (e) L=2, (f) L=1, (g) L=0.

Fig. 7.16 depicts the results of a highly complex image (Fig. 7.16(a)) consisting of extreme illumination variations and diverse textures. As before the GRF, JSEG and EGS fail to handle the variation in illumination conditions, consequently over-segmenting the sky and water regions, as seen in Figs. 7.16 (b), (c) and (d) respectively, which is efficiently dealt with by the proposed algorithm, as seen in Fig. 7.16(g). Moreover since GRF, JSEG and EGS algorithms do not employ a distinct texture descriptor, the textured regions associated with the sea shore, surface of water, components of the boat and other structures in the coastline are subjected to significant over-segmentation. However the MAPGSEG algorithm that employs an entropic texture descriptor is successful in separating regions of disparate texture, as shown in Fig. 7.16(g).

Segmenting textured areas becomes extremely challenging when regions with dissimilar textures have great color similarity. In such a scenario an effective texture descriptor is indispensible to perform the segmentation task. An image of a ‘Cheetah’ has been portrayed in Fig. 7.17(a), whose skin tone complements the color of the background, making it an exceptionally difficult image to segment. The GRF, JSEG, EGS results shown in Figs. 7.17(b), (c), (d) illustrates the effect of an indistinct texture descriptor, while the proposed algorithm in comparison has been able to achieve a good segmentation owed to its distinct texture descriptor, as shown in Fig. 7.17(g). This result emphasizes the importance of the texture characterization module in the MAPGSEG framework for segmentation.

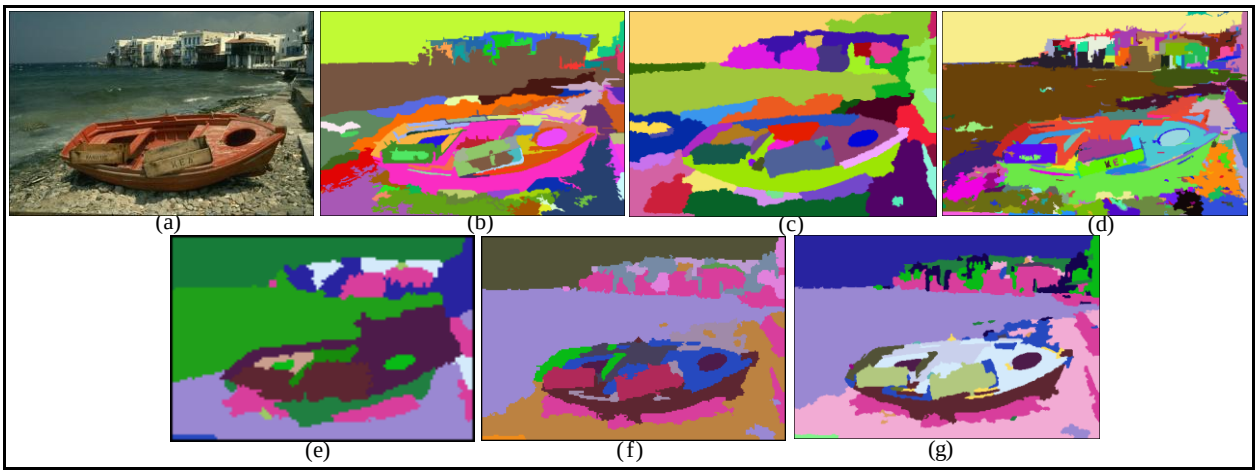

Fig. 7.16: Boat results: (a) original RGB image, (b) GRF, (c) JSEG, (d) EGS, and MAPGSEG (e) L=2, (f) L=1, (g) L=0.

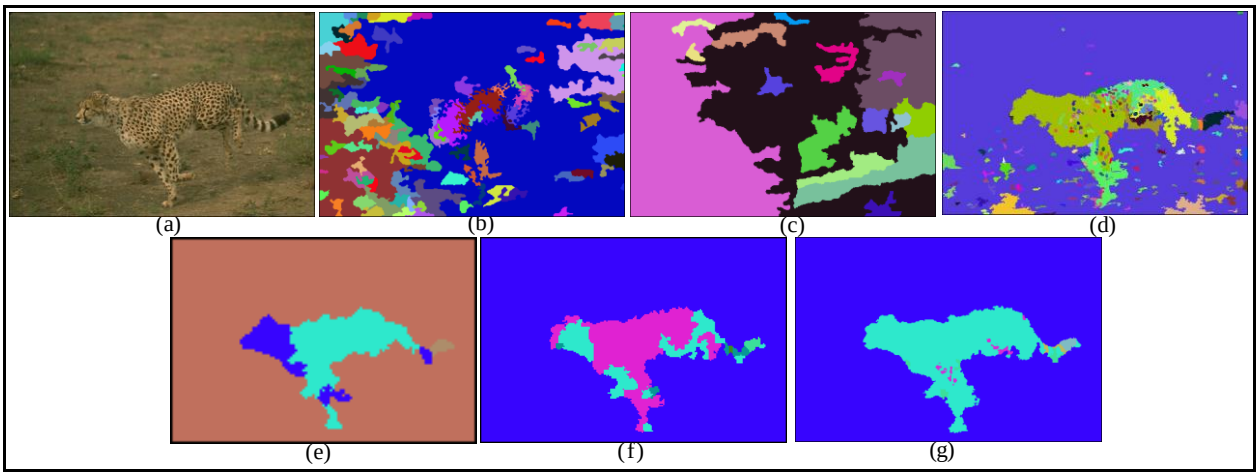

Fig. 7.17: Cheetah results: (a) original RGB image, (b) GRF, (c) JSEG, (d) EGS, and MAPGSEG (e) L=2, (f) L=1, (g) L=0.

Another demanding situation often confronted in segmenting natural scene images is the occlusion of image content by foreground objects, as seen in the 'Race cars' image, depicted in Fig. 7.18(a). In this image it can be observed that the motorway has been partially occluded by the race cars, as a result of which all prior art (Figs. 7.18(b), (c), and (d)) segment the motorway on either side of the vehicles, as two different regions. However the MAPGSEG algorithm which employs the MANOVA-based region merging procedure has the capability of merging spatially independent alike color-texture regions, consequently providing the same label to the motorway on either side of the cars. Thus, the efficiency of the MAPGSEG algorithm in handling the occlusion problem is emphasized in this result. In the Fig. 7.18(g) it can also be observed that the word 'Castrol' has been segmented at multiple locations just about perfectly by the MAPGSEG algorithm, unlike any of the other segmentation methods, which underscores the ability of proposed technique to segment fine details such as text with great competence. For all of

the images displayed so far in this section, the interim segmentation results of the MAPGSEG algorithm at levels *L=2* and *L=1*, have been displayed. The closeness of these interim results to their corresponding final segmentations at level *L=0*, as observed in most cases, signifies the algorithm's robustness to scalability.

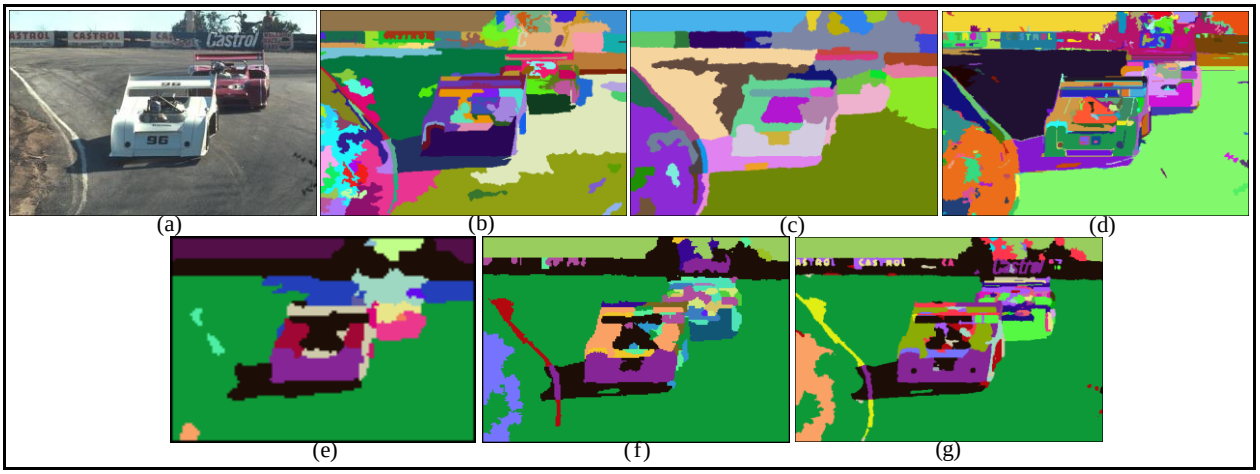

Fig. 7.18: Race cars results: (a) original RGB image, (b) GRF, (c) JSEG, (d) EGS, and MAPGSEG (e) L=2, (f) L=1, (g) L=0.

As mentioned previously, the NPR metric was utilized to evaluate the segmentation correctness of the obtained results. However in order to facilitate the usage of the NPR metric, the evaluation procedure was commenced by computing the expected value of the PR index ($E[PR]$) for all available manual segmentations (described in Appendix B), and was obtained to be 0.6064. A distributional comparison of our evaluation of the segmentation results for 300 images in the Berkeley database, obtained from the GRF, JSEG, EGS, and MAPGSEG is displayed in Figs. 7.19 and 7.20. In this figure, it can be observed that the distribution for the GRF method is weighted more towards the lower half of the distribution with a minimal NPR value as low -0.9. An improvement over the GRF algorithm, are the JSEG and EGS methods where the values are weighted more towards the higher end of NPR score distribution. More favorable NPR scores can be observed in the case of the MAPGSEG. In general the scores that are considerably greater than '0' are considered to be useful, while negative scores imply poor segmentations. Consequently, the actual improvement can be seen observing the number of segmentation scores that fall within the range of very good segmentation results [$0.7<NPR<1$]. These numbers for the GRF, JSEG, EGS and MAPGSEG were computed as 39, 65, 68 and 85 respectively, as shown in Table 7.2. This indicates that approximately a third of the images segmented using our algorithm match closely to the segmentations performed by humans.

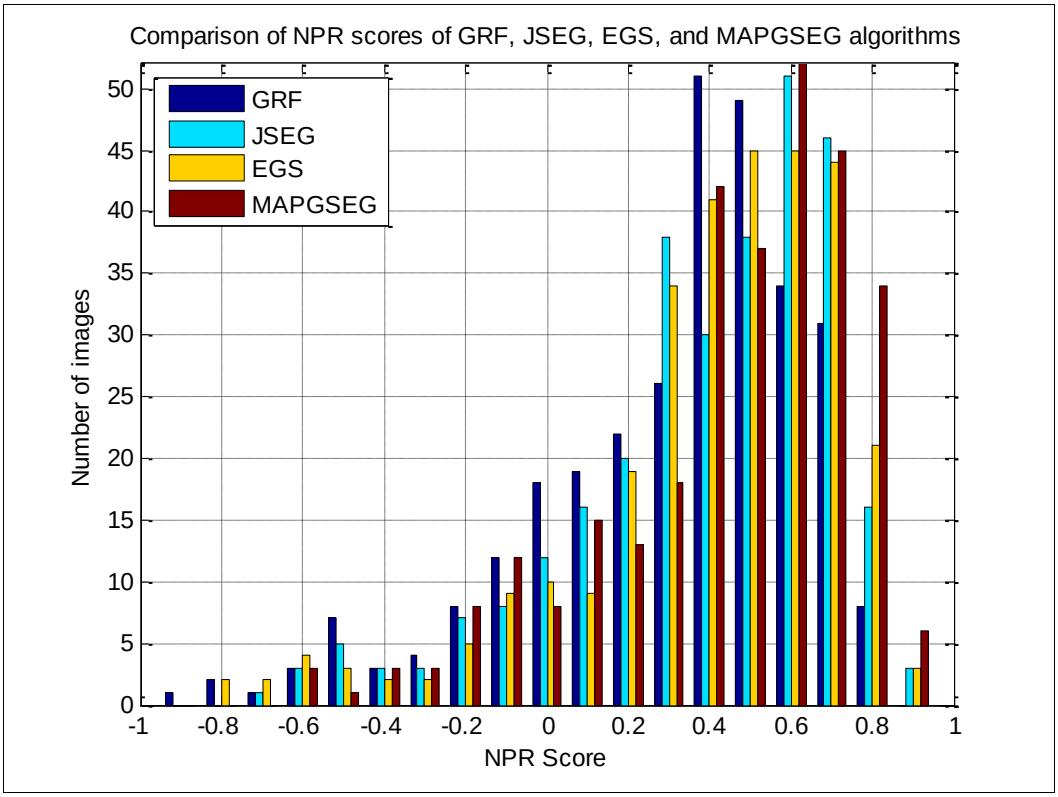

Fig. 7.19: Distribution of NPR scores for 300 images of the Berkeley database

.

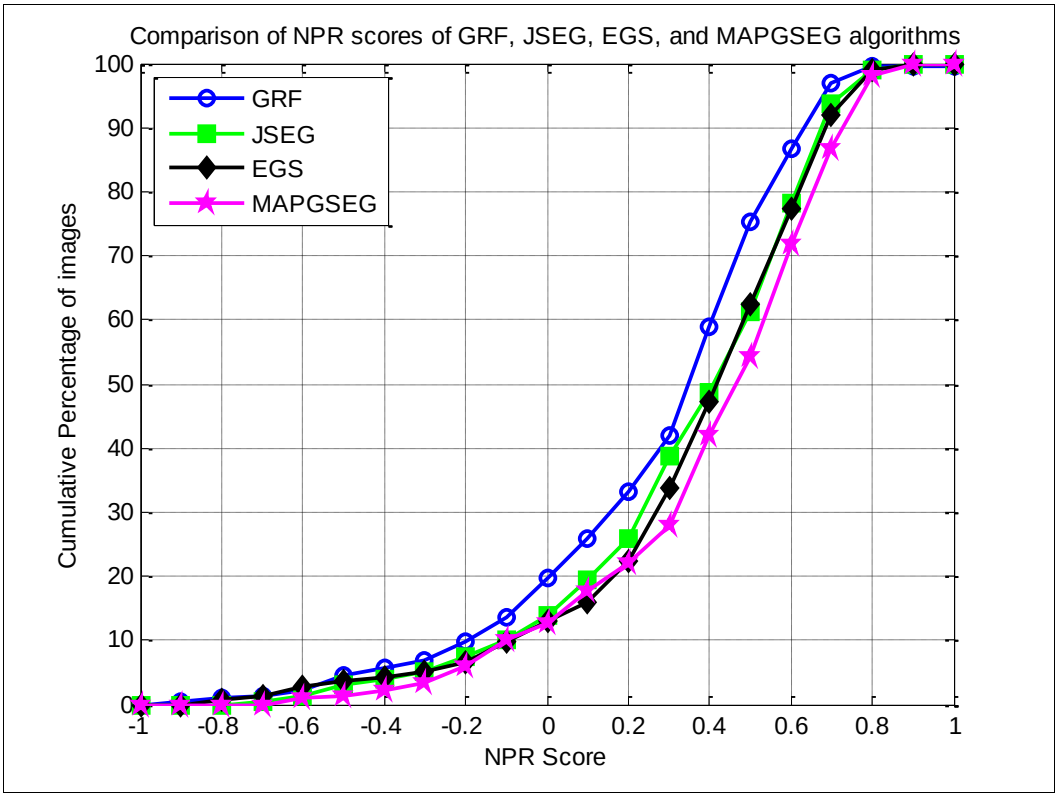

Fig. 7.20: Cumulative Percentage of NPR Scores.

TABLE 7.2: NPR EVALUATION OF SEGMENTATION RESULTS USING 300 IMAGES OF THE BERKELEY DATABASE

	GRF	JSEG	EGS	MAPGSEG
Avg. Time (sec)	240	16.2	7.0	11.1
Avg. NPR	0.358	0.439	0.457	0.495
Std. NPR	0.345	0.318	0.324	0.312
NPR>0.7 (# Images)	39	65	68	85
Environment	C	C	C	MATLAB

A comparison of our evaluation, for the segmentation results obtained from the four methods, is summarized in Table 7.2. This table shows that proposed algorithm has the highest average NPR score, and the narrowest deviation in the scores, signifying that the MAPGSEG algorithm is most consistent in achieving the best segmentations, from a qualitative standpoint. In addition the average run time per image of the MAPGSEG algorithm is comparable to the EGS algorithm which has the lowest average execution time, considering the different environments (C and MATLAB) in which these two algorithms were developed.

Additional results of the MAPGSEG technique in comparison to prior art, under various image scenarios, are shown in Figs. 7.21-7.24. Clear performance advantages of the proposed algorithm discussed in this section can be observed in these results to furnish a fair indication of the algorithm's robustness.

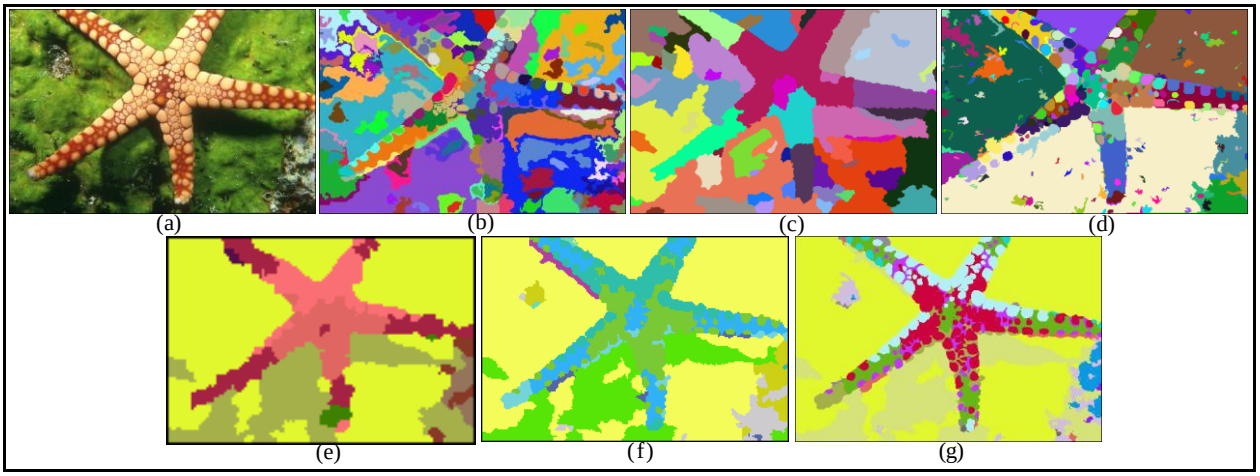

Fig. 7.21: Star fish results: (a) original RGB image, (b) GRF, (c) JSEG, (d) EGS, and MAPGSEG (e) L=2, (f) L=1, (g) L=0.

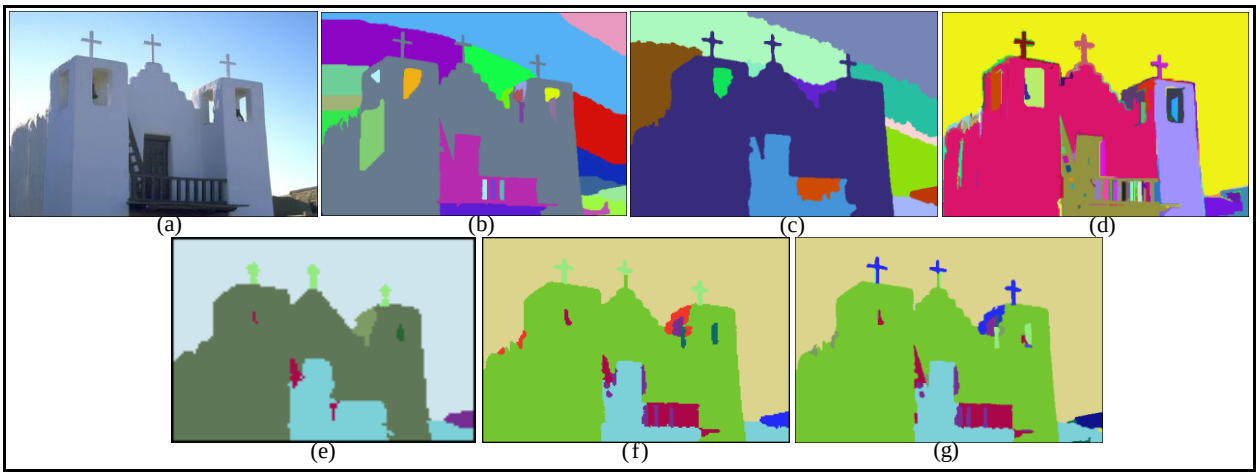

Fig. 7.22: Chapel results: (a) original RGB image, (b) GRF, (c) JSEG, (d) EGS, and MAPGSEG (e) L=2, (f) L=1, (g) L=0.

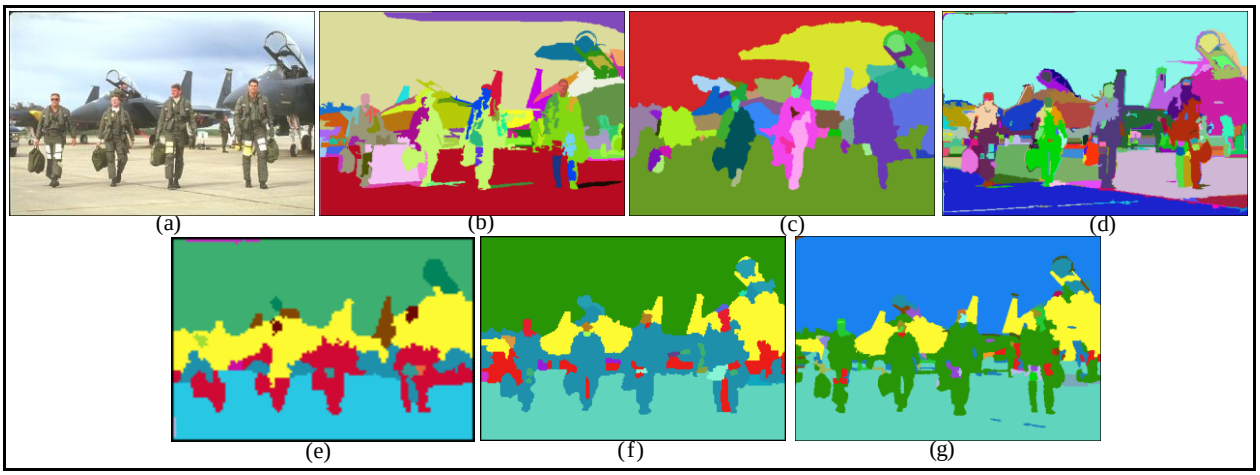

Fig. 7.23: Air force results: (a) original RGB image, (b) GRF, (c) JSEG, (d) EGS, and MAPGSEG (e) L=2, (f) L=1, (g) L=0.

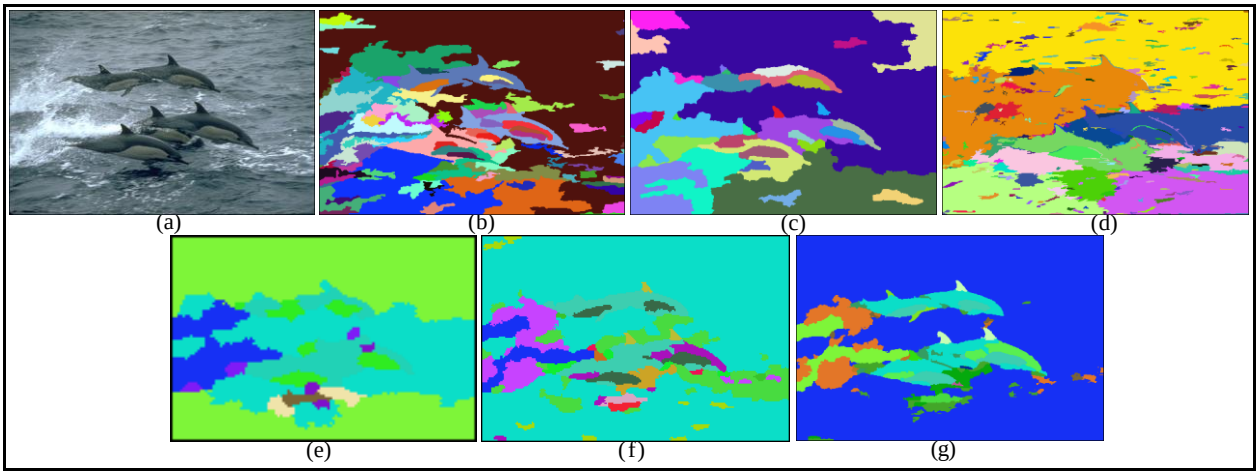

Fig. 7.24: Dolphins results: (a) original RGB image, (b) GRF, (c) JSEG, (d) EGS, and MAPGSEG (e) L=2, (f) L=1, (g) L=0.

7.3. Multi/Hyperspectral Image Segmentation Results

The results of the proposed algorithm were benchmarked on a test-bed of several multi/hyperspectral datasets acquired from a variety of sensors such as HYDICE [Mitchell, 1995], Hyperion [Folkman et al., 2001], Modular Imaging Spectrometer Instrument (MISI) [Feng et al., 1994] and Reflective Optics System Imaging Spectrometer (ROSIS) [Kunkel et al., 1991]. Additionally, our results were compared against two state-of-the-art multi/hyperspectral spatial segmentation algorithms namely, the Size Constrained Region Merging (SCRM) approach developed by Castilla et al. [Castilla et al., 2008] and the Multilevel Hierarchical Image Segmentation (MHIS) method instituted by Li et al. [Li et al., 2011]. The SCRM algorithm was executed with the default parametric settings via ENVI® utilizing an open source binary file provided by Dr. Guillermo Castilla from the Department of Geography, University of Calgary, Canada. On the other hand, the outcomes of the MHIS technique were provided directly by Dr. Peijun Li from the Institute of Remote Sensing and Geographic Information Systems, Peking University, Beijing, China. The results of our algorithm were obtained by using the first few MNF bands spanning more than 99% of the data variability, with S set in the range of 150 to 250 segments depending on the complexity of the scene. Furthermore, our algorithm was developed in MATLAB® and run on a machine having an Intel (R) Core (TM) i7 CPU processor, 2. 79 GHz with 6 GB RAM.

7.3.1 Qualitative Evaluation

The results of the proposed algorithm at interim stages of processing as well as in comparison to the SCRM and MHIS methods are presented in Figs. 7.25-7.28 for four hyperspectral datasets acquired from three different sensors. To this effect, a HYDICE image of an 'Urban' scene with fairly complex spatial content is displayed in Fig. 7.25(a). This image has a mixture of spectral, textural and structural variations, rendering it fairly complex from a segmentation perspective. Fig. 7.25(b) represents a vector gradient map with edge information extracted from the scene. Fig. 7.25(c) shows the initial set of regions acquired at the end of the region growing process. From this figure, it can be seen that this interim region map is over-segmented due to reasons specified in Section 5.5. Figs. 7.25(d)-7.25(f) exhibit respectively three (angular second moment, contrast, and entropy) of the five textural feature maps computed from the first MNF band.

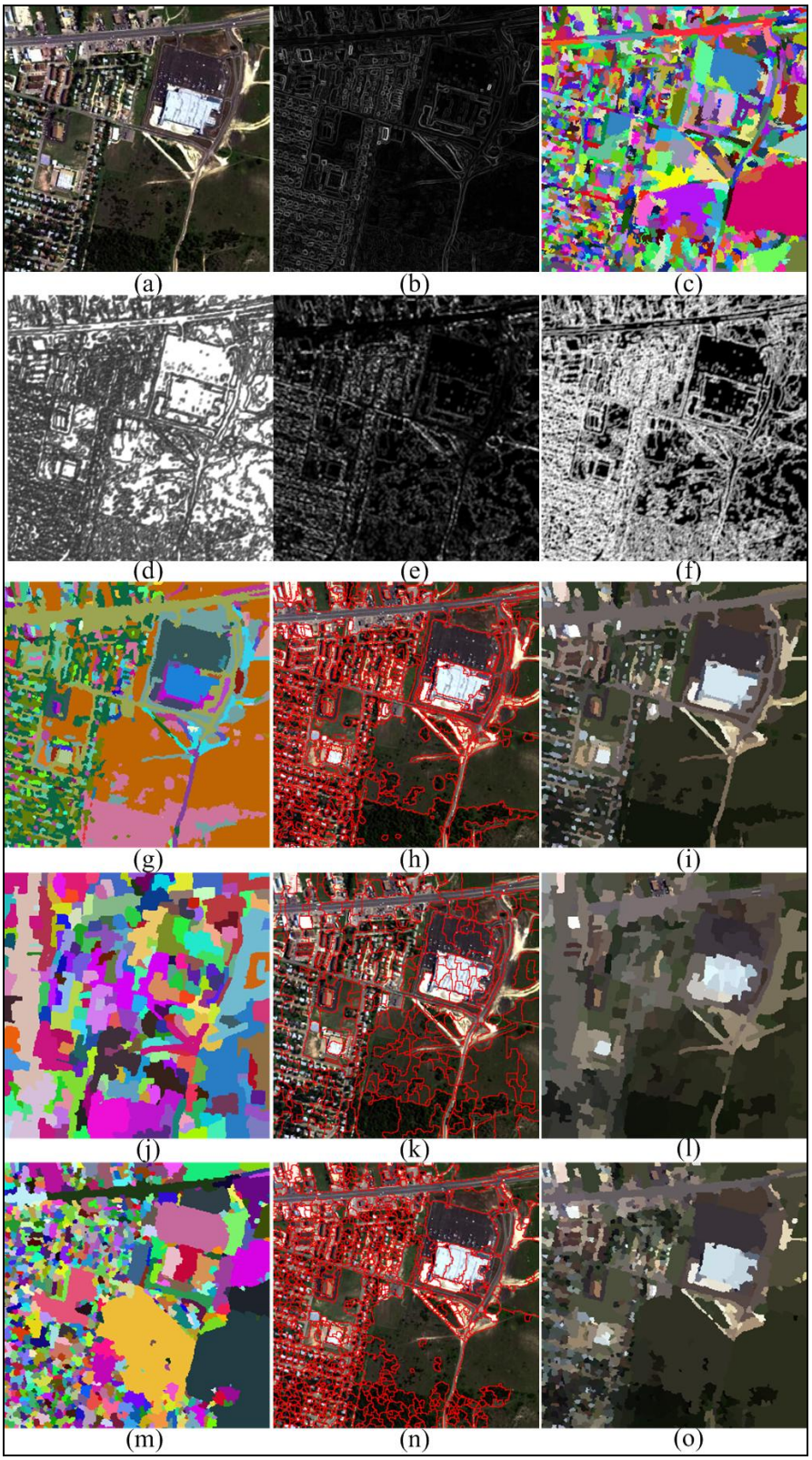

Fig. 7.25: (a) HYDICE 'Urban' scene (bands 49, 35, and 18), (b) Gradient map, (c) Region growth map, (d) f_{ASM}, (e) f_{CON}, (f) f_{ENT}. Segmentation results: (g, h, i) Proposed, (j, k, l) SCRM, and (m, n, o) MHIS.

The output segmentation result acquired at the end of the statistical region merging algorithm is displayed in three alternate representations for ease of viewing (Figs. 7.25(g)-7.25(i)). Specifically, Fig. 7.25(g) portrays the final segmentation as a pseudo color map where each color corresponds to an individual segment. On the other hand, Fig. 7.25(h) depicts a result in which segment boundaries are overlaid on the image, while Fig. 7.25(i) delineates various partitions via a segment-based mean spectral intensity image, in which all pixels within an arbitrary segment are assigned their mean intensity. Finally, results obtained from the SCRM, MHIS methods are shown in Figs. 7.25(j)-7.25(l) and 7.25(m)-7.25(o) respectively. From Figs. 7.25(j)-7.25(l), 7.25(m)-7.25(o), it can be observed that the SCRM and MHIS algorithms suffer from significant over/under-segmentation issues across various parts of the scene as opposed to the proposed scheme. As seen in Figs. 7.25(j)-7.25(l), the SCRM technique over-segments several regions like the building (Wal-Mart), parking lot, highway, tree, and tall/mowed grass areas (all located at the top and right half of the image), while it under-segments houses/residential areas and street components positioned in the left half of the scene. On the other hand, the MHIS methodology facilitates better definition of the Wal-Mart complex (building/parking lot), highway, and grass regions, but produces less optimal segments in the remaining areas pertaining to trees, houses and side-streets, as illustrated in Figs 7.25(m)-7.25(o). The phenomenon of over-segmentation observed with the aforestated mechanisms can be owed to the fact that in these procedures the initial partitions are determined based on a watershed transform that creates numerous small segments in textured areas and in regions with varying shades of spectral information. The under-segmentation caused by the algorithms can be attributed to the absence of a highly localized edge descriptor within their framework to delineate fine details. Conversely, our method utilizes a region growing procedure that simultaneously lays emphasis on the homogeneous and heterogeneous characteristics of a scene via spectral and localized gradient information, in conjunction with multiple textural features in a fusion protocol, to enable the formation of spatially meaningful regions, as seen in Figs. 7.25(g)-7.25(i).

In Fig. 7.26, a HYDICE hyperspectral spectral scene of size 1280 X 307 X 191 imaged over Washington DC and the corresponding segmentation results of a sub-image area (outlined in red) are displayed. This image also presents a challenging setting for segmentation due to the existence of multifarious spatial content involving roads, water bodies, trees, grass, bare soil, buildings to name a few. The results of the SCRM/MHIS routines, in Figs. 7.26(b)-(d) and 7.26(e)-(g) respectively, illustrate their shortfall in furnishing well defined regions that are coherent with the content of the scene, primarily due to the presence of numerous structural and textural discrepancies. However, these variations are more efficiently dealt with by the proposed method yielding a much more satisfactory result. More specifically,

regions associated with grass, trees, water, and concrete/asphalt are distinctly separated in our result unlike the other algorithms.

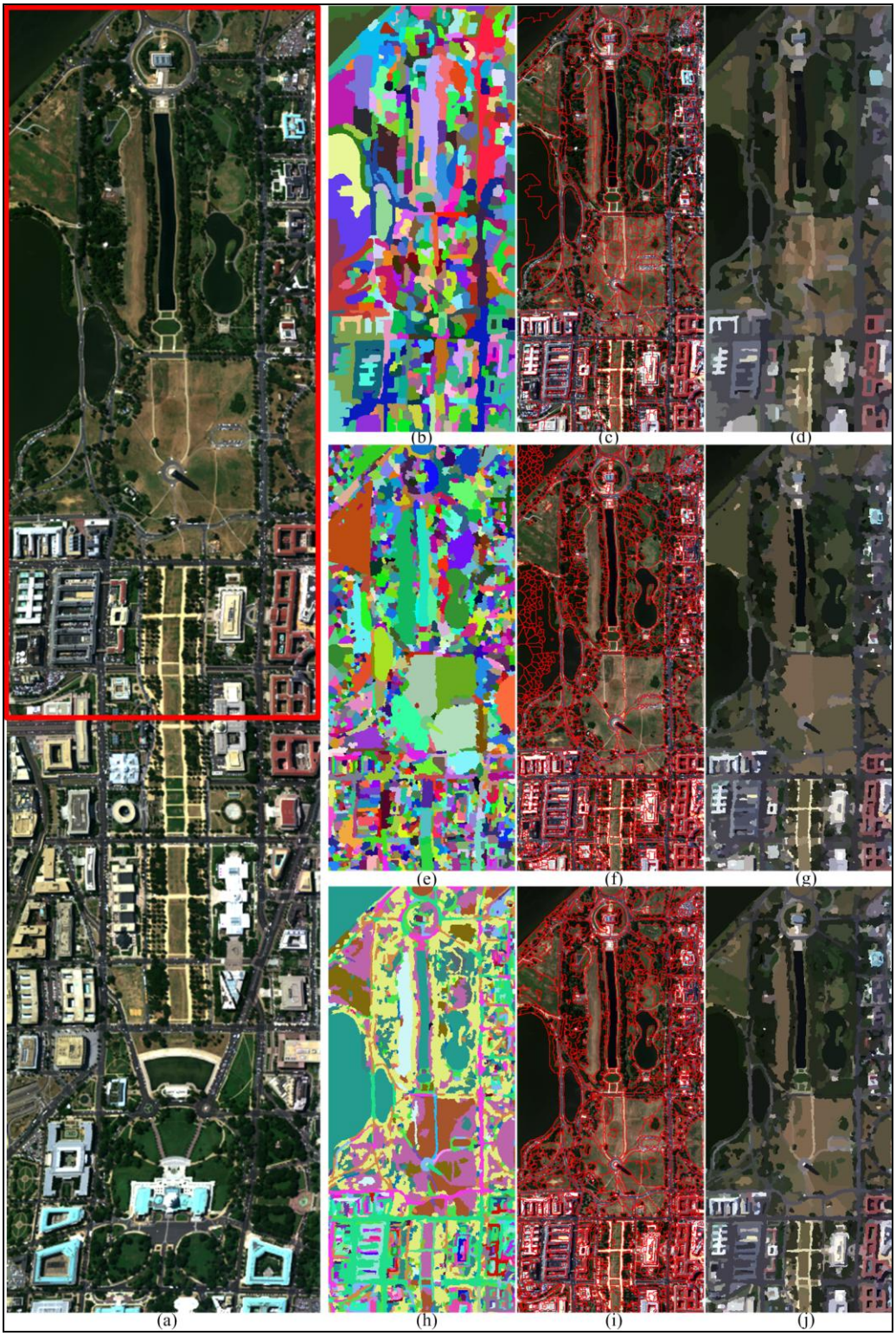

Fig. 7.26: (a) HYDICE 'Washington' scene (bands 49, 35, and 18). Segmentation results: (b, c, d) SCRM, (e, f, g) MHIS, and (h, i, j) Proposed.

Fig. 7.27 shows the results achieved on a hyperspectral dataset (Fig. 7.27(a)) of dimensions 400 X 256 X 196 acquired from the Hyperion sensor imaged over the San Francisco bay. Clear advantages can be discerned in the degree of detail achieved in our segmentation results in comparison to the ones obtained from the SCRM and MHIS approaches. Observe that the bridge, the two big land mass flanking it, as well as the water on either side of the bay area are segmented with high precision by the proposed algorithm as seen in Figs. 7.27(h)-(j). Yet again, the SCRM and MHIS cause over segmentation in these regions due to the reasons specified previously (see Figs. 7.27(b)-(d), and 7.27(e)-(g)).

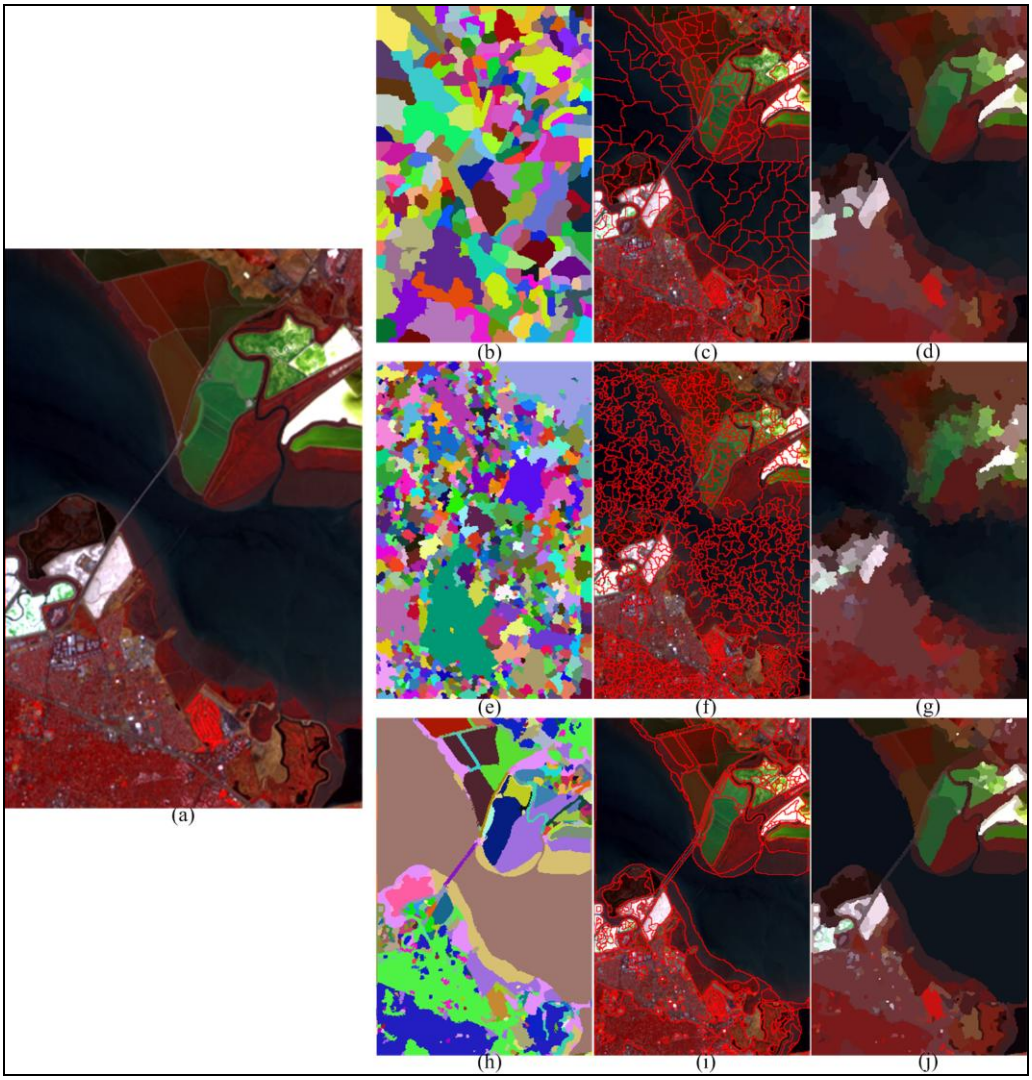

Fig. 7.27: (a) Hyperion 'San Francisco' scene (bands 40, 35, and 18). Segmentation results: (b, c, d) SCRM, (e, f, g) MHIS, and (h, i, j) Proposed.

Segmenting textured regions becomes a harder challenge when regions with diverse textures are similar in spectral composition. In such scenarios, an algorithm equipped with multiple texture descriptors is imperative for region partitioning. Fig. 7.28(a) presents an image of spatial extent 400 X 400 with 61 spectral bands taken from the MISI sensor, of an area dominated by tall trees and small/mowed grass with

comparable spectral composition. Segmenting such an image based solely on their pixel spectra can be an arduous task. The SCRM and MHIS results shown in Figs. 7.28(b)-(d), 7.28(e)-(g) illustrates the effect of the absence of a proper texture descriptor for segmentation. Alternatively, the proposed algorithm (Figs. 7.28(h)-(j)) has been able to achieve an effective segmentation illustrating the robustness of our framework in handling highly textured images as well as emphasizing the importance of the texture characterization module.

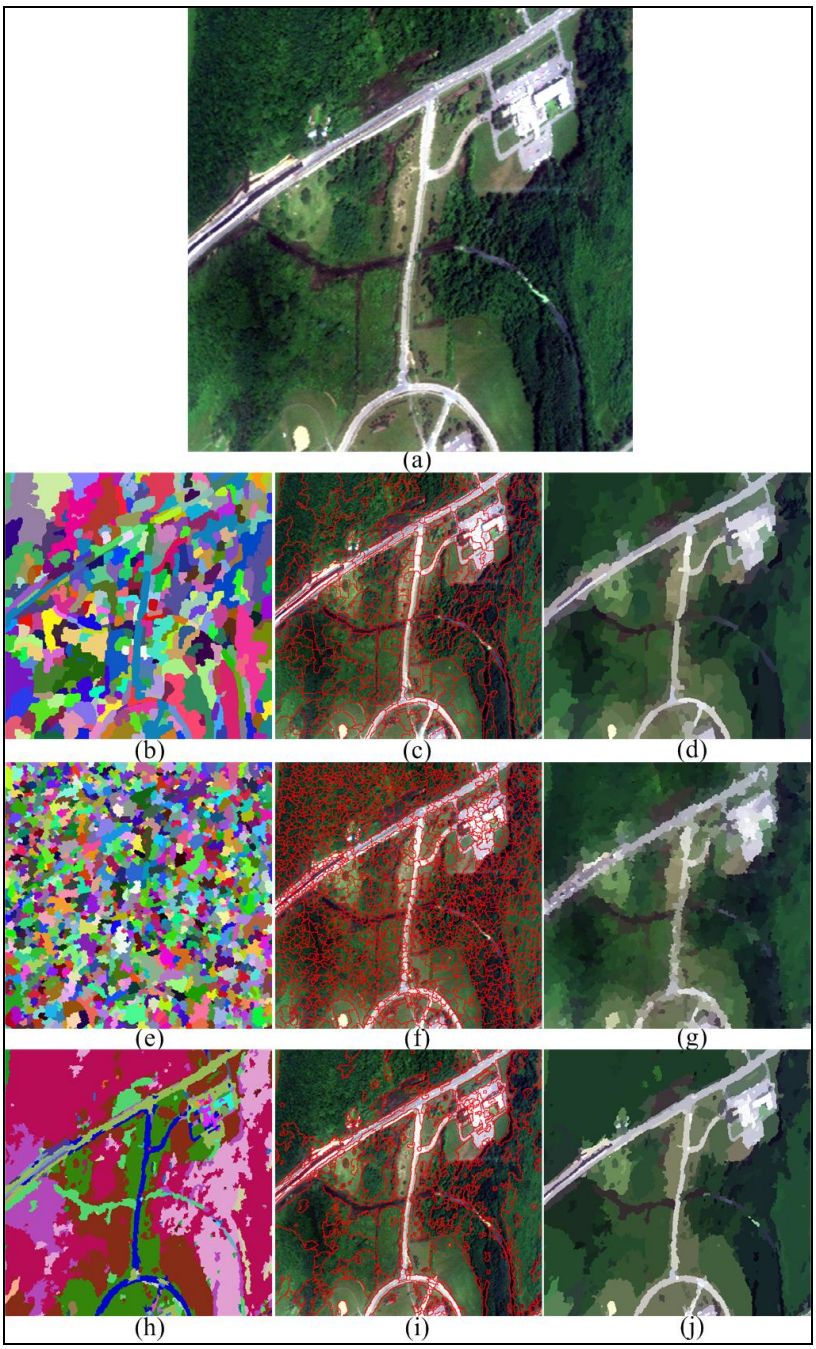

Fig. 7.28: (a) MISI scene (bands 26, 15, and 5). Segmentation results: (b, c, d) SCRM, (e, f, g) MHIS, and (h, i, j) Proposed.

Another aspect of the proposed algorithm demonstrated in the aforementioned set of images Figs. 7.25-7.28 is its capability of handling the "spatial separation" and "occlusion" problems. This is clearly illustrated in Fig. 7.25(g) where multiple tall/mowed grass areas being spatially separated by the highway/side-streets have been segmented out as a single region. A similar observation can be made in Fig. 7.26(h) where several water bodies that are spatially separated across the scene have been assigned the same label by the proposed algorithm. On the other hand, Fig. 7.27 is an example of spatial content occlusion by a foreground object, where the huge water region encompassing the Pacific Ocean and the San Francisco bay has been occluded by a bridge. However the proposed algorithm still manages to uniquely label the two water regions on either side of the bridge. It must be noted that all of the previous scenarios pertinent to spatial separation and occlusion have been efficiently handled due to our unique statistical merging procedure that enables multiple spatially independent yet spectrally and texturally similar regions to be assigned to the same segment label.

7.3.2 Quantitative Evaluation: Segment-based Classification

As applications for segmentation are growing at a swift pace in the remote sensing realm, the development of quantitative performance evaluation methodologies for the obtained segmentation outcomes has had increased attention and is the focus of many ongoing research endeavors. In [Zhang, 1996], a review of approaches that have been instituted to achieve segmentation evaluation from a quantitative standpoint is presented. Most evaluation methods are categorized as being either analytical or empirical in nature. Analytical methods evaluate segmentations by analyzing the principles and properties of the algorithms that generated them. Conversely, empirical techniques indirectly quantify the quality of segmentations of an arbitrary algorithm by using specific test images that have reference/ground truth results, to compute goodness and/or discrepancy metrics between them.

In order to objectively evaluate the quality of our segmentation outcomes, we adopt an empirical/indirect evaluation technique involving analysis of segmentation via classification [Linden et al., 1996]. The underlying hypothesis is that if a segmentation result comprises of partitions that perfectly align with objects/uniform regions in a scene, then the mean spectral signatures computed from such segments will distinctly represent individual classes. This not only avoids confusion between them but also produce class maps that are more homogeneous in nature. On the other hand, for an under/over-segmentation result where segments are larger or smaller than meaningful entities in a scene, the mean spectral values of such segments will span across multiple classes (if under-segmented) or misrepresent them (if over-segmented), resulting in sub-optimal classification due to the increased confusion among classes.

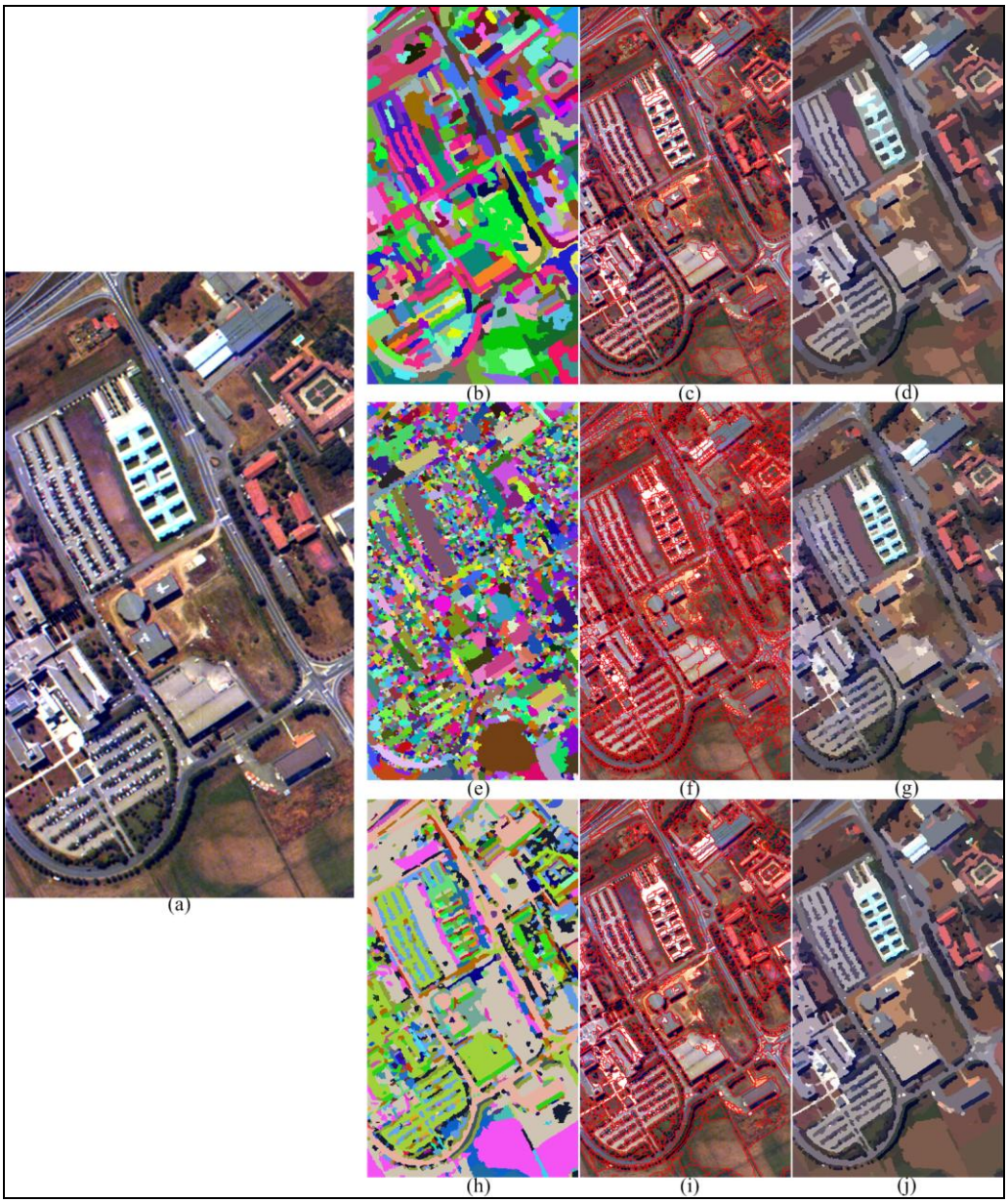

Fig. 7.29: (a) ROSIS 'University' scene (bands 60, 30, and 2). Segmentation results: (b, c, d) SCRM, (e, f, g) MHIS, and (h, i, j) Proposed.

Results of the proposed algorithm were quantitatively analyzed using two hyperspectral data sets named 'University' and 'Center', both acquired from the Reflective Optics System Imaging Spectrometer (ROSIS) sensor. These datasets along with ground truth training and testing samples for nine class-types have been provided by Dr. Paolo Gamba from the University of Pavia, Italy. Furthermore, the Gaussian Maximum Likelihood (GML) and Support Vector Machine (SVM)-based classification algorithms were used in the evaluation. While the GML algorithm was implemented in MATLAB®, SVM-based classification was performed using a multi-class pairwise SVM classifier inherent in the LIBSVM library [Chang and Lin, 2011], with a Gaussian radial basis function kernel. The images in Fig. 7.29 showcase

the segmentation results of the 'University' scene (Fig. 7.29(a)) with spatial-spectral extent of 640 X 310 X 103, obtained from the SCRM (Figs. 7.29(b)-(d)), MHIS (Figs. 7.29(e)-(g)) and proposed algorithm (Figs. 7.29(h)-(j)). It imperative to understand that once the segmentation maps (Figs. 7.29(b, e, h)) from various algorithms were procured, corresponding segment-based mean spectral signature images were generated (Figs. 7.29(d, g, j) and utilized as inputs to the GML and SVM classification algorithms. Fig. 7.30 shows manually selected regions of interest (ROIs) for nine classes- asphalt, meadows, gravel, trees, metal sheets, bare soil, bitumen, self-blocking bricks and shadow, acquired to train (left) the two classifiers as well as test (right) their classification accuracy.

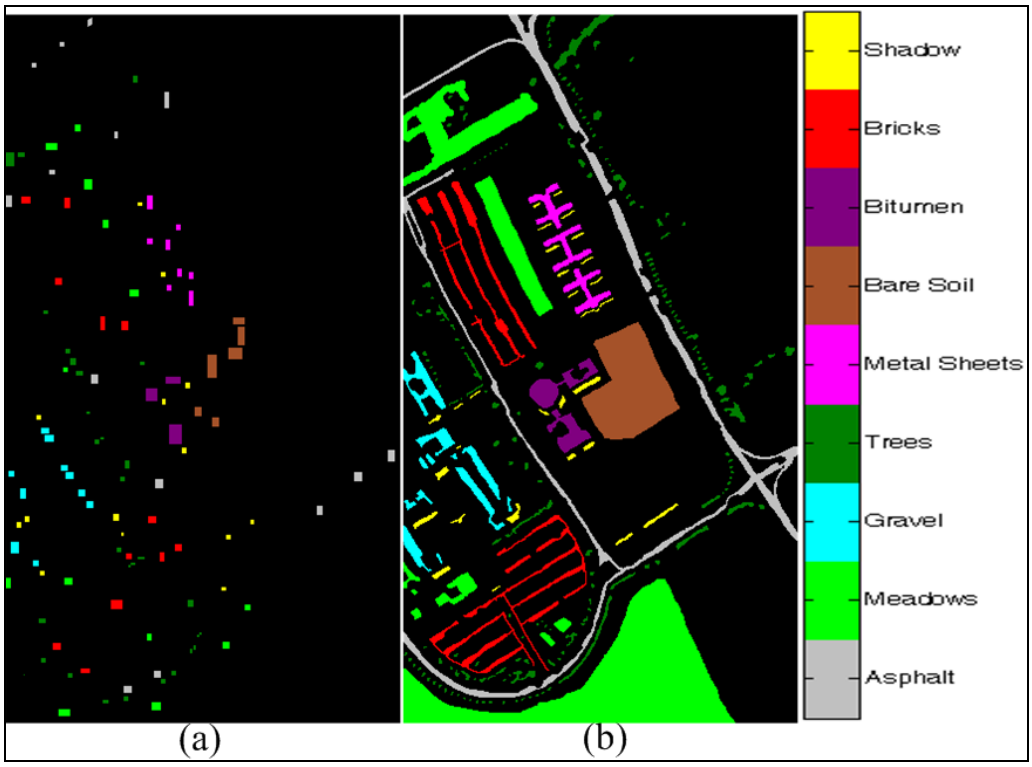

Fig. 7.30: ROSIS 'University' scene: (a) Training, and (b) Testing samples.

The results of the GML, SVM classifiers for the 'University' dataset in pixel and segment-based architectures are illustrated in Fig. 7.31. Figs. 7.31(a) and 7.31(e) represent pixel-based classification maps from the aforestated classifiers obtained using the first few MNF bands that account for more than 99% of the variability in the data. Additionally, Figs. 7.31(b)-(d), 7.31 (f)-(h) respectively portray the GML and SVM segment-based classification results, where the segments were outlined by the SCRM, MHIS and the proposed algorithm. Figs. 7.32-7.34 are the direct counterparts of Figs. 7.29-7.31, exhibiting the segmentation and classification results on a 1096 X 301 X 102 sized sub-image extracted from a hyperspectral ROSIS scene titled 'Center', imaged over the city center in Pavia, Italy. The classification was done with respect to nine classes- water, trees, meadows, bricks, base, soil, asphalt, bitumen, tiles and shadow, as depicted in Fig. 7.33.

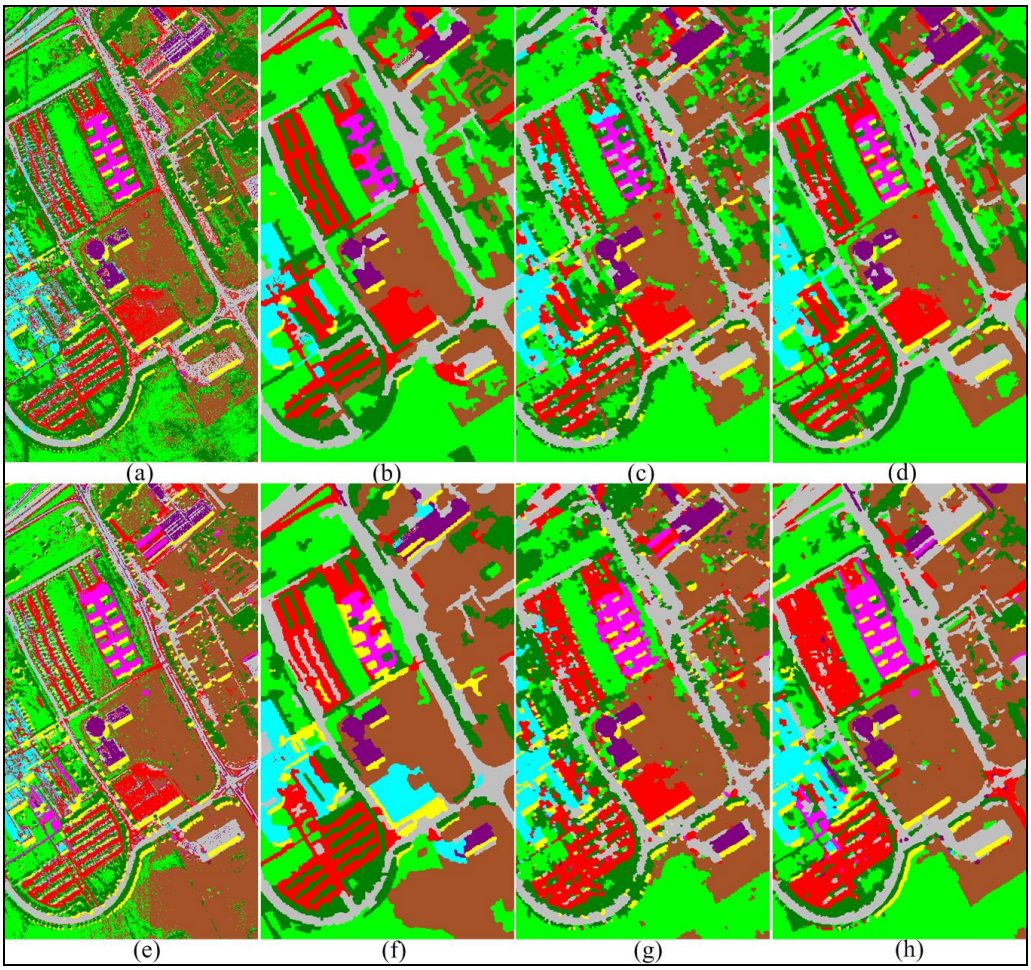

Fig. 7.31: GML classification results: (a) Pixel-based, (b) SCRM, (c) MHIS, and (d) Proposed. SVM classification results: (e) Pixel-based, (f) SCRM, (g) MHIS, and (h) Proposed

Fig. 7.32: (a) ROSIS ‘Center’ scene (bands 60, 30, and 2). Segmentation results: (b, c, d) SCRM, (e, f, g) MHIS, and (h, i, j) Proposed.

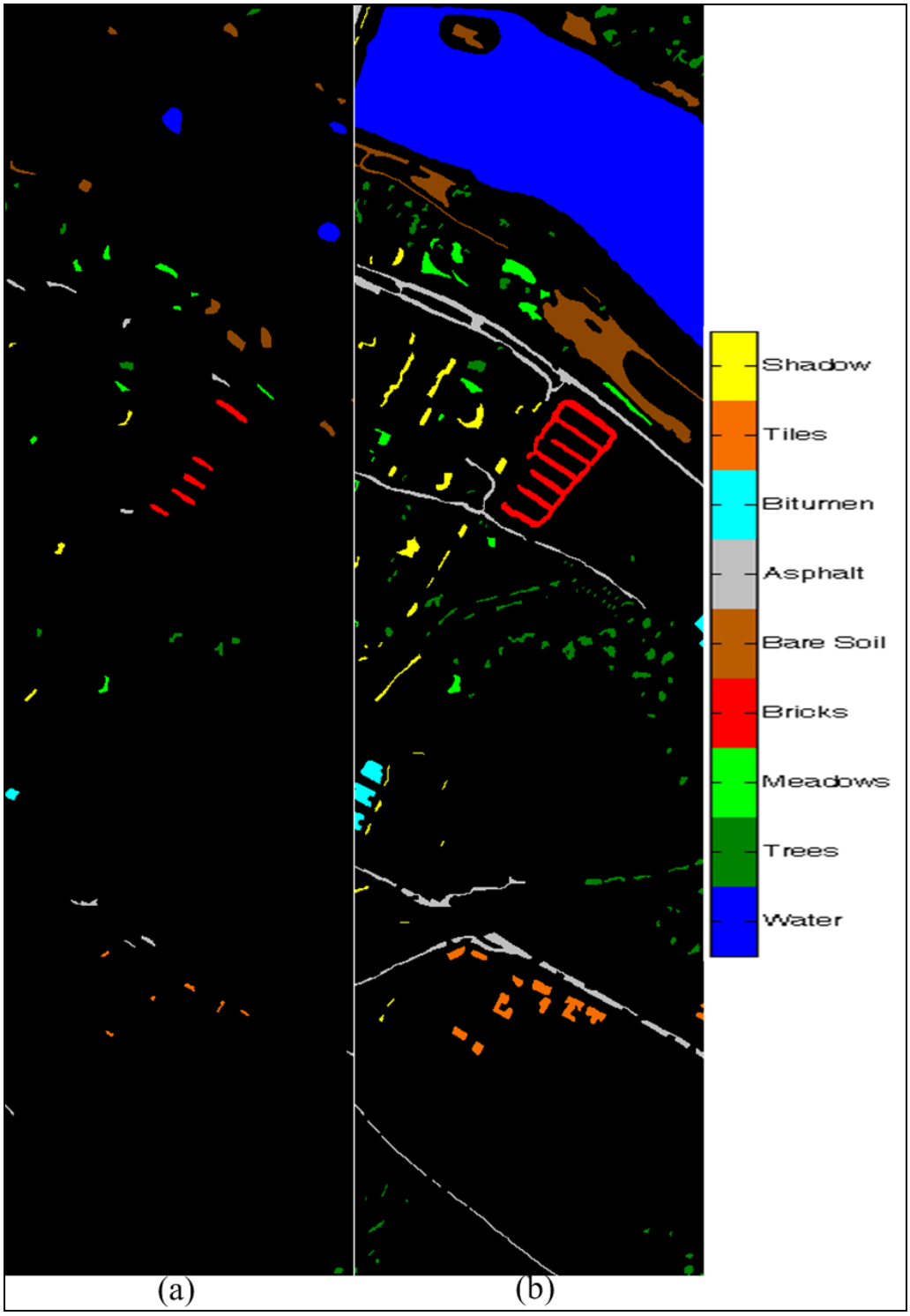

Fig. 7.33: ROSIS 'Center' scene: (a) Training, and (b) Testing samples.

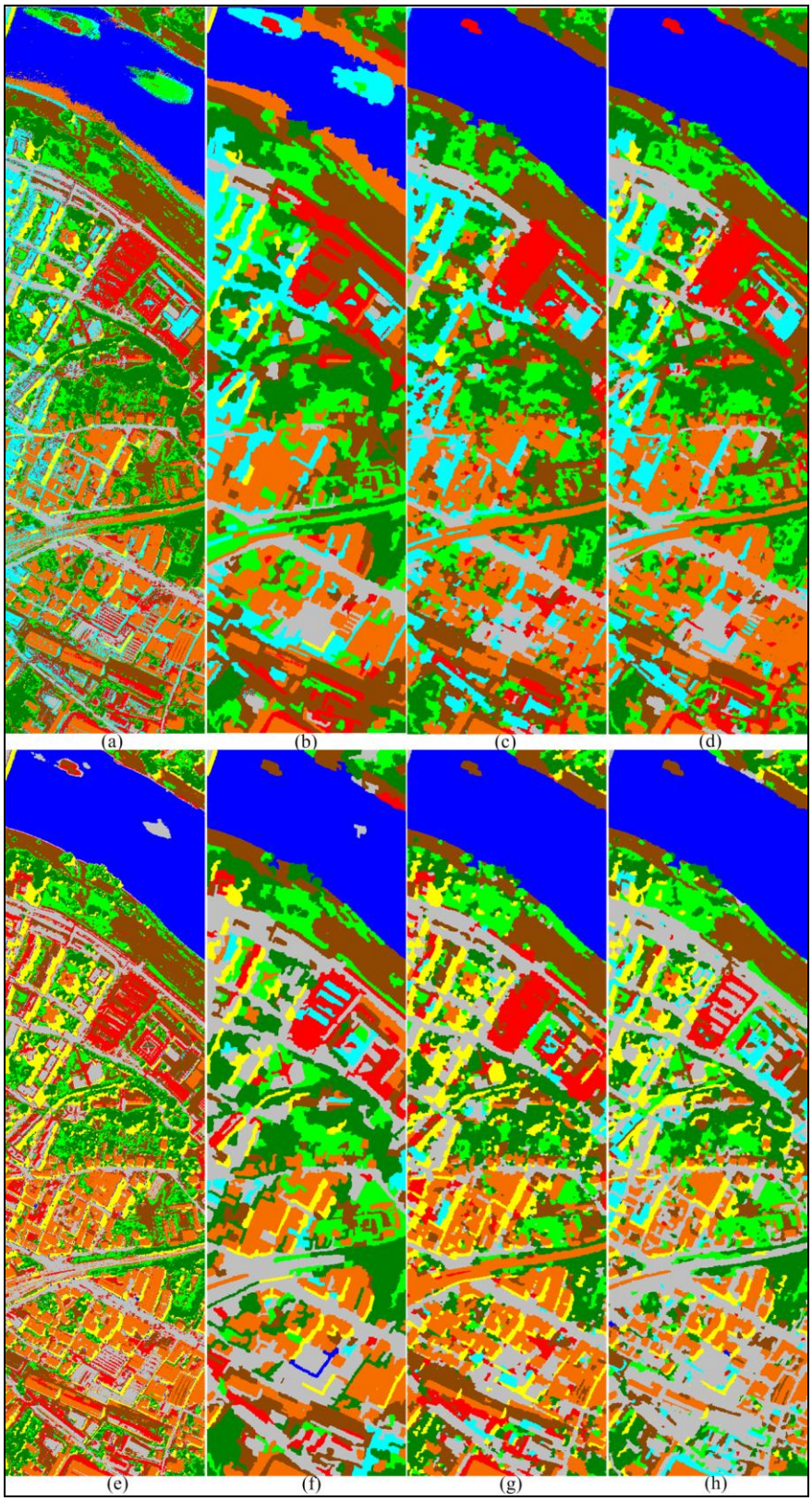

Fig. 7.34: GML classification results: (a) Pixel-based, (b) SCRM, (c) MHIS, and (d) Proposed. SVM classification results: (e) Pixel-based, (f) SCRM, (g) MHIS, and (h) Proposed.

In order to numerically evaluate the class maps displayed in Fig. 7.31 and Fig. 7.34, the test ROIs for the two images (Figs. 7.30(b) and 7.33(b)) were utilized to compute the accuracy of the GML and SVM classifiers for the various classes, tabulated in Tables 7.3 and 7.4. Furthermore, the overall accuracy (OA), average accuracy (AA), Kappa coefficient and the overall misclassification (Mis) rate [Tarabalka et al., 2010] across all classes were calculated (summarized in Tables 7.5 and 7.6) to determine the most accurate class maps. From Tables 7.5 and 7.6, it can be seen that for both datasets, the segment-based class maps acquired from our algorithm has the highest overall percentage accuracy and Kappa coefficient, with the lowest misclassification rate, illustrating that the spatial segments generated using the proposed algorithm had the best agreement with the boundaries of various objects/uniform regions leading to class-maps consistently better than the other algorithms (including the conventional approach of pixel-based classification). Moreover, the results in Figs. 7.31 and 7.34 clearly corroborate the notion that in general segment-based class maps are much more homogeneous relative to pixel based classification.

TABLE 7.3: INDIVIDUAL CLASSIFICATION ACCURACIES (%) FOR THE UNIVERSITY IMAGE

	Samples		GML				SVM			
	Train	Test	Pixel	SCRM	MHIS	Proposed	Pixel	SCRM	MHIS	Proposed
Asphalt	548	6631	81.18	86.22	81.44	**92.79**	75.74	88.73	87.08	**88.96**
Meadows	540	18649	74.21	97.47	99.60	**99.76**	53.42	62.09	91.47	**97.55**
Gravel	392	2099	**76.70**	61.70	63.17	64.70	64.65	98.76	77.32	**99.67**
Trees	524	3064	**98.30**	71.90	90.05	85.44	**94.84**	82.90	85.15	90.63
Metal Sheets	265	1345	**100.00**	90.33	98.74	99.26	99.70	90.33	98.07	**99.93**
Bare Soil	532	5029	91.65	98.85	96.38	**99.26**	95.65	**99.36**	99.20	97.91
Bitumen	375	1330	92.26	86.02	96.69	**97.67**	88.80	99.47	**100.00**	99.40
Bricks	514	3682	85.17	73.19	75.10	**97.37**	92.61	98.10	90.30	**98.32**
Shadow	231	947	**98.31**	57.13	60.93	88.91	**99.89**	82.47	96.73	93.35

TABLE 7.4: INDIVIDUAL CLASSIFICATION ACCURACIES (%) FOR THE CENTER IMAGE

	Samples		GML				SVM			
	Train	Test	Pixel	SCRM	MHIS	Proposed	Pixel	SCRM	MHIS	Proposed
Water	615	42477	85.99	76.22	**100.00**	**100.00**	98.07	99.49	**100.00**	**100.00**
Trees	334	2998	**94.80**	86.96	77.95	89.56	86.06	84.16	**87.96**	83.89
Meadows	378	1457	**99.04**	73.78	92.52	98.01	97.32	74.06	89.70	**98.22**
Bricks	485	2140	90.47	18.97	99.30	**99.39**	68.50	**97.10**	77.57	92.62
Bare Soil	780	5603	**96.56**	89.31	94.47	95.36	96.16	93.65	98.25	**99.25**
Asphalt	372	4847	**94.53**	50.86	64.51	80.63	96.78	**97.59**	96.20	97.32
Bitumen	80	640	81.09	**85.94**	**85.94**	**85.94**	0.31	39.22	16.25	**98.13**
Tiles	120	1193	**99.66**	94.38	93.80	91.79	98.99	94.38	**99.66**	91.03
Shadow	138	1665	**96.04**	66.43	23.18	61.68	**99.88**	75.74	97.84	95.38

TABLE 7.5: GLOBAL CLASSIFICATION ACCURACIES (%) FOR THE UNIVERSITY IMAGE

	GML				SVM			
	Pixel	SCRM	MHIS	Proposed	Pixel	SCRM	MHIS	Proposed
OA	82.039	88.739	90.8523	**95.3502**	72.3186	79.4932	91.0394	**95.9744**
AA	88.6427	80.3116	84.6767	**91.685**	85.0332	89.1361	91.7022	**96.19**
Kappa	77.3813	84.9792	87.8483	**93.8105**	66.2889	74.5829	88.3085	**94.6974**
Mis	17.961	11.261	9.1477	**4.6498**	27.6814	20.5068	8.9606	**4.0256**

TABLE 7.6: GLOBAL CLASSIFICATION ACCURACIES (%) FOR THE CENTER IMAGE

	GML				SVM			
	Pixel	SCRM	MHIS	Proposed	Pixel	SCRM	MHIS	Proposed
OA	88.93	74.13	93.24	**96.22**	95.28	96.09	97.07	**98.36**
AA	**93.13**	71.43	81.30	89.15	82.45	83.93	84.83	**95.09**
Kappa	81.19	59.12	87.20	**92.84**	91.17	92.59	94.42	**96.88**
Mis	11.07	25.87	6.76	**3.78**	4.71	3.91	2.93	**1.64**

7.4. 3-D Segmentation Results of Medical Volumes

The results of the proposed algorithm were evaluated on multiple MRI/CT scans provided by the Massachusetts General Hospital (MGH), National Biomedical Imaging Archive (NBIA), and the DataPhysics Research (DPR) Incorporation. Specifically, the unsupervised version of the proposed algorithm (discussed in Section II) were qualitatively and quantitatively evaluated on 20 normal MR brain data sets and their manual segmentations provided by the Center for Morphometric Analysis at the MGH (available online at http://www.cma.mgh.harvard.edu/ibsr/). The MRI datasets were of spatial dimensions 256 x 256 while the number of slices varied from 60 to 64. On the other hand, the supervised segmentation methodology (illustrated in Section III) was demonstrated using two different CT studies on 3 target sub-volumes (lungs, spleen, and vertebral structure) provided by the NBIA and DPR. The datasets were of spatial dimensions 512 x 512 while the number of slices varied from 100 to 150. Additionally, our results were compared against a state-of-the-art 3-D active contour-based segmentation algorithm developed by Yushkevich et al. [Yushkevich et al., 2006] made available as an open source software tool called ITK-SNAP. The ITK-SNAP algorithm was executed utilizing a downloadable file provided by Dr. Paul Yushkevich from the Penn Image Computing and Science Laboratory (PICSL) in the Department of Radiology at the University of Pennsylvania. The results of our algorithm were run from MATLAB® files developed at the Rochester Institute of Technology, with S set equal to15 segments (for the automatic segmentation framework) depending on the complexity of the volume. Furthermore, our algorithm was run on a machine having an Intel (R) Core (TM) i7 CPU processor, 2. 79 GHz with 6 GB RAM.

7.4.1 Qualitative Evaluation

The results of the proposed algorithm at interim stages of processing as well as in comparison to the ITK-SNAP method are presented in Figs. 7.35-7.39 for different MRI/CT datasets.

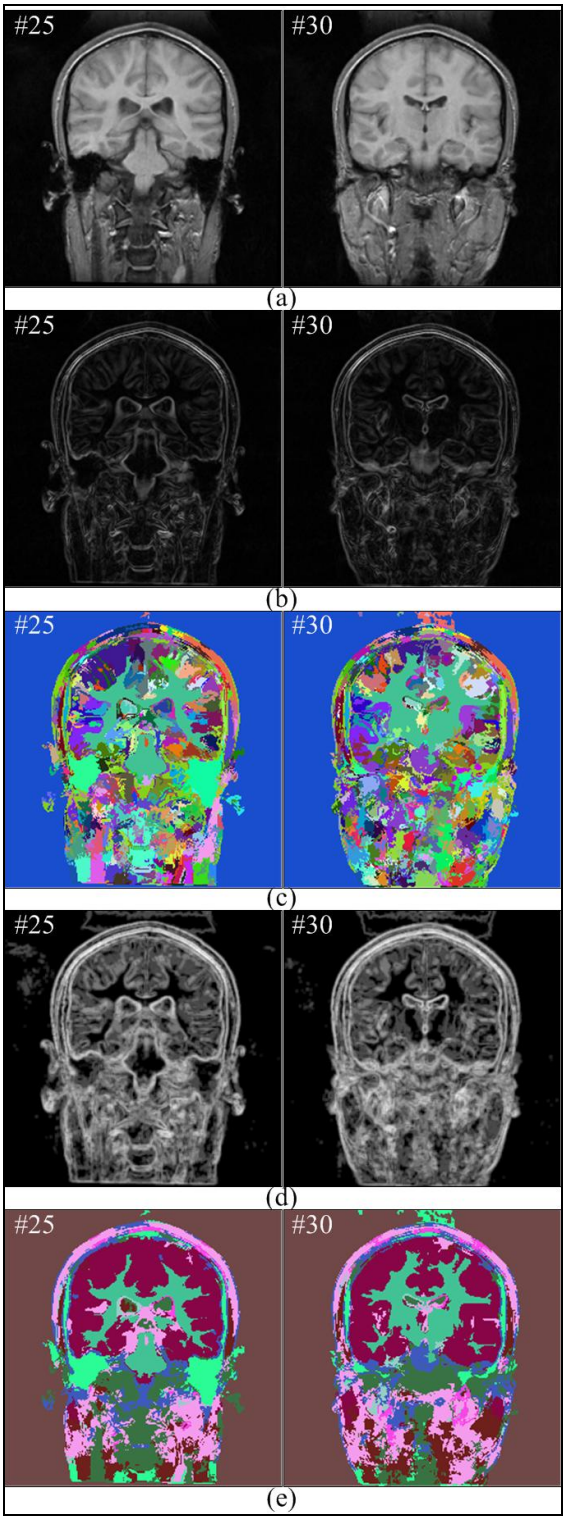

Fig. 7.35: (a) Slice #25 and 30 of an MRI volume "11_3", with corresponding (b) Gradient maps, (c) Volume growth maps, (d) Texture maps, and (e) final segmentation outcomes.

Fig. 7.35(a) shows two slices (#25 and 30) from a sample MRI volume named "11_3" (11 and 3 are the patient identification and scan number respectively) of dimensions 256 x 256 x 63. Fig. 7.35(b) represents the corresponding gradient maps acquired using the 3-D vector field gradient estimation technique discussed in Section II A. Fig. 7.35(c) shows the initial set of sub-volumes acquired at the end of the volume growing process. It is important to note that the output of the proposed algorithm is a unique label assignment for every volumetric entity (in a 3-D matrix) each associated with a randomly generated pseudo color for visual distinction, as seen in Fig. 7.35(c). From this figure, it can be seen that set of sub-volumes formed at the conclusion of the growing process are over-segmented due to reasons specified in Section II D. Fig. 7.35(d) exhibits the outcomes of the 3-D entropy-based texture descriptor computed from the input volume. The output segmentation result acquired at the end of the MANOVA-based volume merging algorithm is displayed in Fig. 7.35(e). As, mentioned previously, Fig. 7.35(e) portrays the final segmentation as a pseudo color map where each color corresponds to an individual segment.

A big challenge with segmenting MRI data is the segregation of gray-matter and white-matter areas of the volume due to their intricate layout in the brain. Gray-matter is a major component of the brain that predominantly comprises of neuronal cell bodies and is positioned in the peripheral areas of the brain. Conversely, the white-matter section contains myelinated axon tracts and is location in the central portion of the brain engulfed within the gray-matter. In order to visually analyze the performance of the proposed approach against the ITK-SNAP algorithm, we extracted the partitions associated with the gray and white- matter sections of the input volume from the automatically segmented result of the proposed algorithm and rendered them as 3-D entities in Fig. 7.36(a). In this figure, the sub-volume on the left (rendered in red) is the gray- matter while, the sub-volume on the right (reproduced in green) is the white-matter. Fig. 7.36(b) shows the corresponding 3-D views of the gray and white- matter sub-volumes derived using the ITK-SNAP algorithm as described in [Yushkevich et al., 2006], while Fig. 7.36(c) portrays the reference/ground truth segmentations for these areas provided by the MGH, performed manually using the techniques described in [Kennedy et al., 1989] and [Filipek et al., 1994]. From Fig. 7.36(b), it can be observed that the ITK-SNAP algorithm suffers from significant errors in partitioning gray- matter, while comprises of much less detail in the segmented outcome of the white- matter, in comparison to the proposed scheme. The aforestated phenomenon are observed due to the fact that in case of geodesic active contour-based procedures such as ITK-SNAP, the obtained partitions are formed at the expense of a few errors and ignoring minor image features for efficiently minimizing the cost function being utilized. Conversely, our method utilizes a 3-D volume region growing procedure that simultaneously lays emphasis on the homogeneous and heterogeneous characteristics of the data via intensity and 3-D gradient information, in conjunction with a 3-D entropy-based texture descriptor in an information fusion mechanism (3-D volume merging), enables the formation of volumetrically compact

gray and white- matter sub-volumes, as seen in Figs. 7.36(a). Moreover from Fig. 7.36 it can be observed that the results of the proposed algorithm are visually in much closer proximity to the ground truths than the ITK-SNAP outputs.

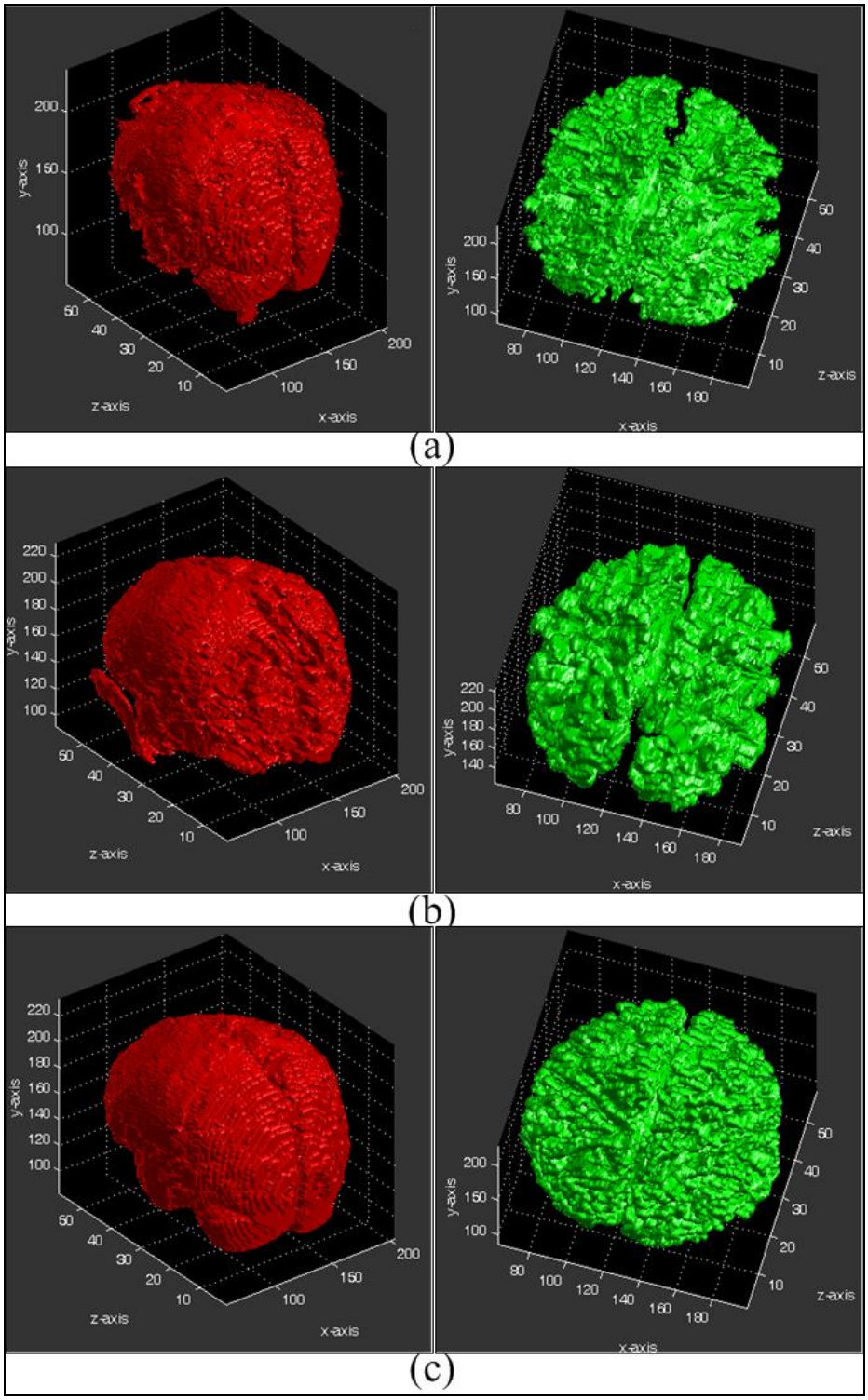

Fig. 7.36: Gray/white-matter segmentation results: (a) ITK-SNAP (b) Proposed algorithm, and (c) Ground Truth segmentations.

As opposed to the performing segmentation in a fully- automated process as discussed in Fig. 7.35 – 7.36 for MRI volumes, Figs. 7.37-7.39 demonstrate the potential of the our segmentation framework for selective segmentation of three target sub-volumes lungs, spleen and vertebral structure in CT data, in line with the work flow of modern day radiologists.

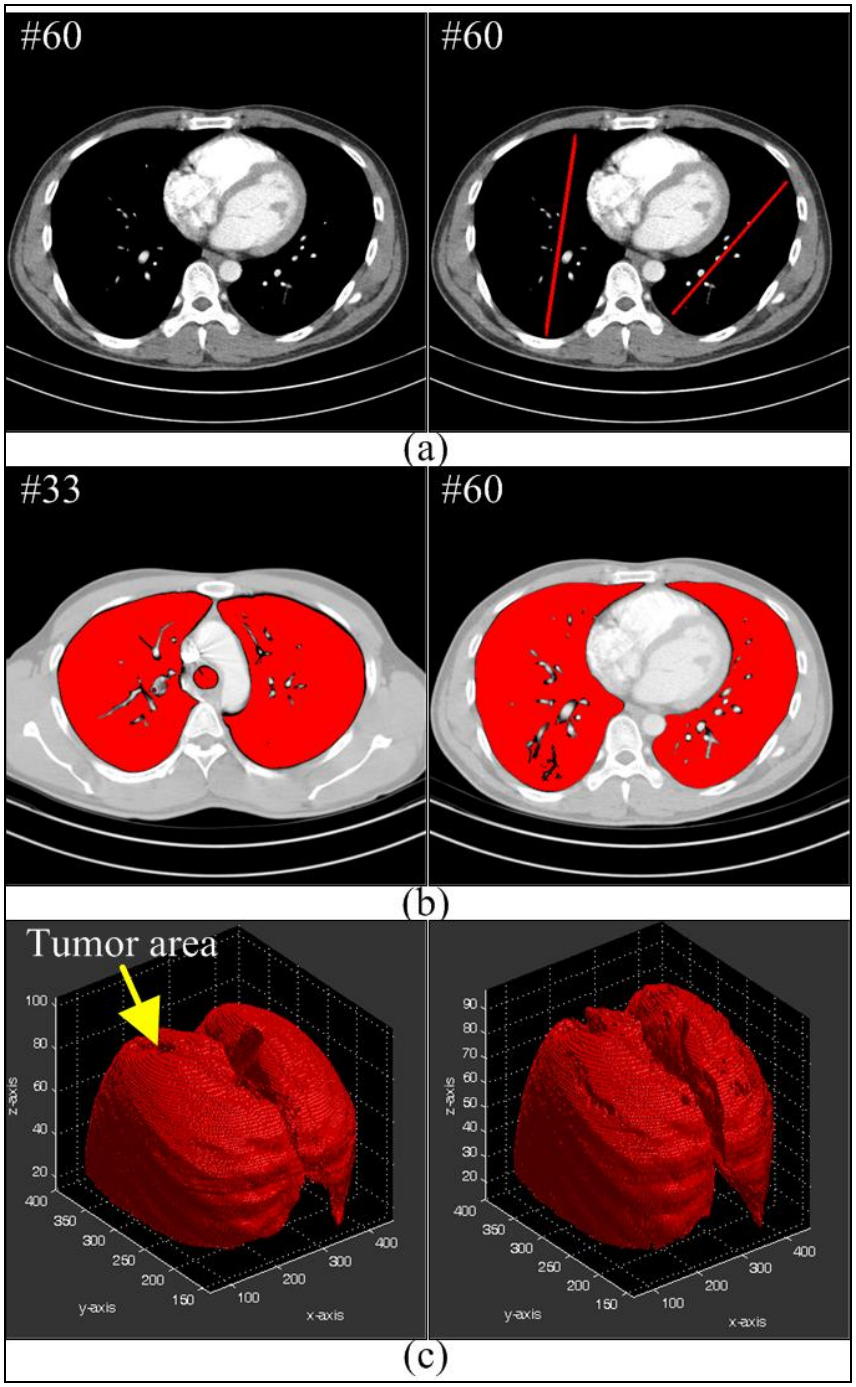

Fig. 7.37: Target sub-volume - Lungs: (a) Arbitrary CT slice (left) and user-defined line markings in the SVOI (right), (b) 2-D view of the corresponding segmentation using the proposed algorithm across two different slices, and (c) 3-D view of the segmentation results using the proposed method (left) and ITK-SNAP (right).

Abstraction of the lungs is a relatively complex task due to the presence of numerous nodules contained within them. In Fig. 7.37(a), on the left is shown is an arbitrary slice (#60) obtained from a sample study of dimensions 512x512x104, publicly available at the NBIA. The right image of Fig. 7.37(a) depicts two user-defined line markings made in the lungs. Fig. 7.37(b) illustrates the planar view of the segmented area (in red) across two arbitrarily chosen slices (#33 and # 60) of the same volume. The aforementioned images illustrate the competence of our method for lung segmentation, by avoiding grouping of large nodules as being part of them. In Fig. 7.37(c), shown are the segmented lungs in the 3-D using our framework (left) and the ITK-SNAP algorithm (right). This result illustrates the shortfall of the ITK-SNAP algorithm in furnishing well defined lung regions in the vicinity of numerous structural variations (like the ribs) encompassing them, as well as a tumor rested on top of the left lung. Conversely, these complex surroundings have been efficiently dealt with using the proposed algorithm. The proficiency of the proposed algorithm can also be observed in Fig. 7.38 wherein the segmentation outcomes of the spleen acquired using a sample study of dimensions 512x512x136, provided by the DPR Inc.

Segmentation of the vertebral structure in the abdomen is a difficult task due to its complex structure. In Fig. 7.39(a) shown are 16 user initializations (on the right) demarcated within the cross-section of the bones in slice #29 of the previously mentioned study used for spleen segmentation. Fig. 7.39(b) furnishes the 2-D view of the segmented abdominal bone area (red) for two slices (#29 and #100) of the volume. Finally, Fig. 7.39(c) details the 3-D segmentation results using the proposed algorithm on the left, and the ITK-SNAP algorithm on the right. The results in Fig. 7.39(b) – 7.39(c) demonstrate the potential of the proposed method to segment complex bone structures pertinent to the ribs, spine and the pelvic girdle. On the other hand, the ITK-SNAP algorithm in spite of using several contour initializations across multiple different slices fails to segment the entire bone structure.

Overall, Figs. 7.35-7.39 not only indicate our algorithm's efficacy in achieving desirable segmentations, with competing qualitative performance against a state-of-the art technique described in [Yushkevich et al., 2006], but also demonstrate the generalization potential of proposed algorithm across different modalities such as MRI/CT for the segmentation task in both an automated as well as semi-automated architecture (as preferred by radiologists). Additionally, in ITK-SNAP framework several different parameters (eg. number of balloons and their size, algorithm iterations, location and number of slices in which balloon should be placed etc., to name a few) and stringent convergence criteria have to be set for each run of the algorithm, which is not the case with the proposed algorithm where in the automated version at most S may be changed based on the complexity of the volume.

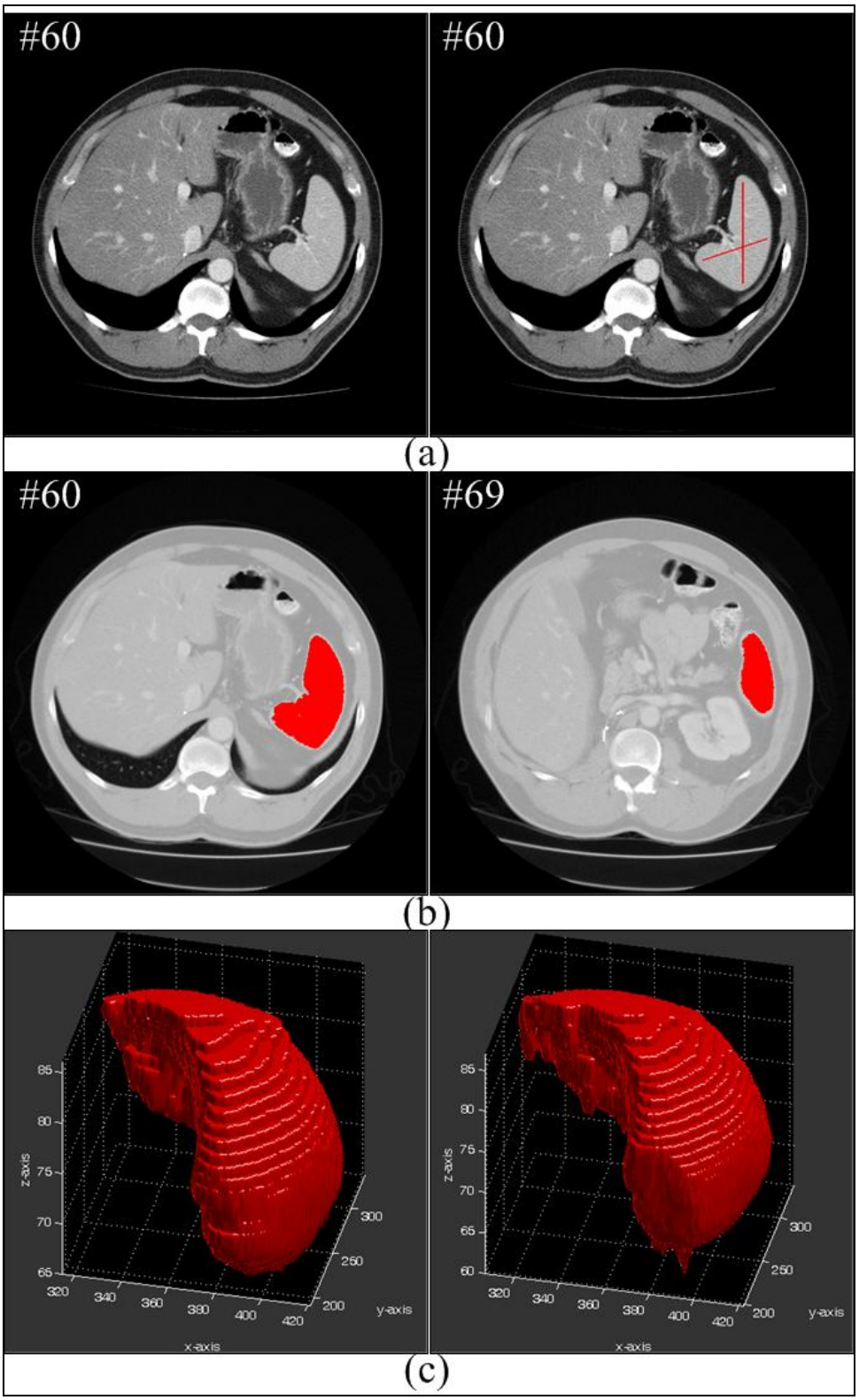

Fig. 7.38: Target sub-volume - Spleen: (a) Arbitrary CT slice (left) and user-defined line markings in the SVOI (right), (b) 2-D view of the corresponding segmentation using the proposed algorithm across two different slices, and (c) 3-D view of the segmentation results using the proposed method (left) and ITK-SNAP (right).

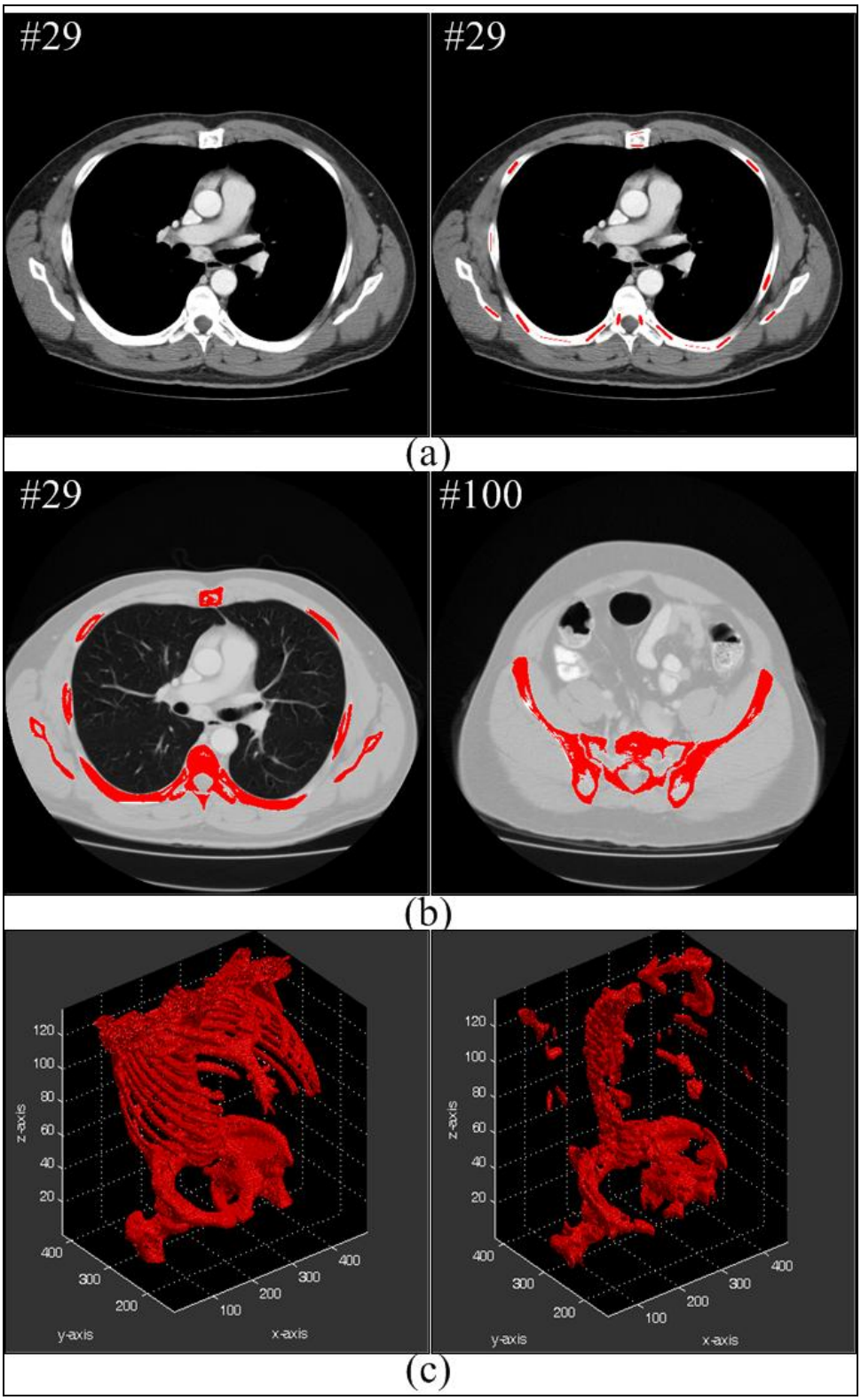

Fig. 7.39: Target sub-volume – Vertebral structure: (a) Arbitrary CT slice (left) and user-defined line markings in the SVOI (right), (b) 2-D view of the corresponding segmentation using the proposed algorithm across two different slices, and (c) 3-D view of the segmentation results using the proposed method (left) and ITK-SNAP (right).

7.4.2 Quantitative Evaluation

Given the availability of 20 normal MR brain data sets and their corresponding manual segmentations for gray/white-matter areas provided by the MGH, we quantitatively evaluated the segmentation performance of the proposed algorithm (unsupervised framework) against ITK-SNAP in an empirical fashion. More specifically, the overall accuracy (OA) - percentage of correctly segmented pixels, and error rate (ER) - percentage of erroneously segmented pixels, for both categories across all 20 datasets were calculated, as depicted in Tables 7.7 and 7.8.

TABLE 7.7: GRAY- MATTER SEGMENTATION ACCURACIES (%)

	Proposed		**ITK-SNAP**	
Patient ID _ Scan #	**OA**	**Mis**	**OA**	**Mis**
100_23	94.43	5.57	89.64	7.36
110_3	95.57	4.43	89.45	7.55
111_2	94.60	5.40	90.58	6.42
112_2	96.30	3.70	91.47	5.53
11_3	95.29	4.71	90.79	6.21
12_3	94.79	5.21	91.16	5.84
13_3	94.86	5.14	90.31	6.69
15_3	89.10	10.90	90.79	6.21
16_3	95.68	4.32	91.11	5.89
17_3	93.99	6.01	91.33	5.67
191_3	96.31	3.69	91.83	5.17
1_24	95.95	4.05	91.03	5.97
202_3	95.50	4.50	91.80	5.20
205_3	96.04	3.96	92.22	4.78
2_4	91.23	8.77	90.78	6.22
4_8	93.10	6.90	90.08	6.92
5_8	93.52	6.48	90.35	6.65
6_10	91.52	8.48	89.77	7.23
7_8	88.13	11.87	90.72	6.28
8_4	95.68	4.32	91.12	5.88
Mean	**94.08**	**5.92**	**90.82**	**6.18**

TABLE 7.8: WHITE- MATTER SEGMENTATION ACCURACIES (%)

	Proposed		ITK-SNAP	
Patient ID _ Scan #	**OA**	**Mis**	**OA**	**Mis**
100_23	94.39	2.61	92.83	4.17
110_3	94.07	2.93	92.20	4.80
111_2	93.57	3.43	92.30	4.70
112_2	94.71	2.29	93.17	3.83
11_3	94.12	2.88	92.83	4.17
12_3	93.74	3.26	93.10	3.90
13_3	93.51	3.49	92.67	4.33
15_3	94.24	2.76	92.87	4.13
16_3	94.65	2.35	92.85	4.15
17_3	93.99	3.01	92.88	4.12
191_3	94.87	2.13	93.46	3.54
1_24	94.52	2.48	92.64	4.36
202_3	94.52	2.48	93.25	3.75
205_3	94.63	2.37	93.49	3.51
2_4	94.80	2.20	92.40	4.60
4_8	93.60	3.40	91.23	5.77
5_8	91.01	5.99	92.74	4.26
6_10	94.01	2.99	91.28	5.72
7_8	95.08	1.92	93.37	3.63
8_4	94.93	2.07	92.78	4.22
Mean	**94.15**	**2.85**	**92.72**	**4.28**

From these tables, it can be seen that for both gray-matter as well as white-matter, the segmentation outputs acquired from our algorithm has the highest overall percentage accuracy with the lowest error rate, illustrating that the outcomes generated using our approach has better agreement with the reference segmentations.

Chapter 8

Summary of Accomplishments, Conclusions and Potential Directions of Future Research

This chapter provides a high level summary of the research accomplishments achieved in this doctoral study, along with few conclusions as well as anticipated future work.

8.1 Summary of Accomplishments

In Chapter 3, an effective method for automatic segmentation of color images (first research objective) called GSEG designed to handle images of varying complexity, has been presented. The algorithm is primarily based on procedures involving color-edge detection, dynamic region growth, entropy based texture characterization and culminates in a unique multivariate region merging technique. The algorithm has been qualitatively and quantitatively tested on a large database of images including the publicly available Berkeley segmentation database, and the quality of results show that our algorithm is robust to various image scenarios and is superior to the results obtained on the same image when segmented by several popular methods, as can been seen in the results shown.

This doctoral work also presents a computationally efficient method designed for fast unsupervised segmentation of color images in a multiresolution framework (second research objective), described in Chapter 4. The proposed Multiresolution Adaptive and Progressive Gradient-based color image SEGmentation (MAPGSEG) algorithm primarily constitutes edge detection, progressive region growth involving distributed dynamic seed addition, multiresolution seed transfer and concludes in a multi-attribute region merging procedure. Performance evaluation of our results (similar to the GSEG algorithm) on several hundred images utilizing a recently proposed metric called the Normalized Probabilistic Rand (NPR) index demonstrates that the proposed algorithm computationally outperforms published segmentation techniques with superior quality.

Also presented in this work is an efficient approach for segmenting multi/hyperspectral imagery (third research objective) into spatially meaningful regions (detailed in Chapter 5), to help reduce the search space in the spatial domain for applications such as target/anomaly/change detection, and 3-D scene reconstruction to name a few. The proposed algorithm was qualitatively benchmarked against two state-of-the-art segmentation approaches for remotely sensed imagery and the results favorable performance in the presence of spectral, textural and structural disparities as well as scenarios encompassing spatial separation, and occlusion.

Finally, Chapter 6 provides an illustration of a novel multi-attribute 3-D medical volume segmentation technique in automated and semi-automated frameworks, for applications such as 3-D reconstruction, visualization and navigation. Our segmentation framework was benchmarked against the ITK-SNAP method with favorable qualitative and quantitative performance over multiple MRI/CT volumes. Additionally, we discussed a generic mathematical framework for effective gradient detection in 3-D, with the potential of accommodating 3-D vector field data.

8.2 Conclusions

Segmentation continues to be at the forefront of many commercial and research endeavors, and the need for algorithms that perform this task efficiently is exponentially increasing with no sign of subsiding in the near future. Modern segmentation approaches (like the ones developed in this doctoral work) have successfully managed to achieve high levels of sophistication and quality, due to increased efforts to develop algorithms that combine the strengths of multiple processes to overcome existing drawbacks.
This observation can be corroborated through the results shown achieved in Chapter 7.

Furthermore, enormous technological strides have been made in the capture/representation of data derived from multiple modalities of imaging, as different modalities provide different levels and type of information. However, segmentation techniques that successfully leverage information manifested in these enhanced formats are relatively few in comparison. To this effect, this doctoral work demonstrates a segmentation framework developed with a multi-modal perspective, to simultaneously cater to the requirements of sophisticated applications that employ datasets acquired from diverse imaging modalities like remote sensing and medical imaging whilst being backward compatible with applications using conventional gray scale/color imagery.

8.3 Potential Directions of Future Research

8.3.1 Digital / Full Motion Video Segmentation

Content analysis, analytics, manipulation, compression and storage of full motion video (FMV) has been growing at an unprecedented rate for numerous commercial and military applications. From a commercial standpoint, FMV content acquired from conventional video cameras, camcorders are being uploaded and watched at a torrential rate via user-friendly websites such as YouTube. Additionally, FMV data is also being acquired by the military via airborne sensors and UAVs for extracting intelligence to perform day-to-day reconnaissance, combat support, forensic analysis, security, and search/rescue duties. Hence, several FMV Terabytes are being uploaded daily, and manually analyzed, contributing to a multi-billion dollar budget. In general, these FMV databases contain multi-dimensional information spanning a variety of spatial (texture, gradient, thermal), spectral, and temporal (motion) attributes that require critical and

custom handling by analysts to achieve optimum visual/application/system performance. Most often, customers or analysts are interested in quickly and effectively understanding/interpreting spatial and/or temporal content segments to achieve a given objective or provide accurate intelligence. For instance, one may be interested in: 1) adjusting the fleshtone or memory colors in spatiotemporal data (3-dimensional rendering), 2) retrieving a specific event in a FMV sequence (like suspicious vehicle movement down a highway), 3) spatiotemporally performing a forward predictive analysis or backward forensic analysis of events/targets, and/or 4) compressing and storing FMV content in a flexible and effective manner. While applications like these are growing at a swift pace in the commercial/military/geospatial intelligence realms, the rapid technological advances being made in the capture and sharing of FMVs has not only resulted in large repositories of information but also simultaneously created a huge discrepancy in the rate at which they are collected versus manipulated or analyzed. This has compelled a pertinent need for the development of effective and efficient automated systems/tools for faster analysis of FMVs.

Given the above challenges, the research done in this doctoral study can be leveraged to: 1) establish a firm foundation for high levels of analysis, exploitation, interpretation, enhancement, classification, storage and compression of FMVs by performing meaningful spatiotemporal region/object/event/shot segmentation, 2) efficiently combine the abundance of the spectral, gradient, motion and textural information in FMVs, for identification and localization of desired content to extract intelligence information, 3) explore the potential of FMV segmentation as a 3-D spatiotemporal problem as opposed to a conventional 2-D frame-by-frame correspondence/estimation problem, and 4) facilitate effective segmentation for objects/regions that are similar but spatially separated and/or undergoing varying degrees of occlusion. Formally, FMV segmentation is defined as the meaningful partitioning of frames into non-overlapping spatiotemporally homogeneous or nearly homogeneous regions/objects from a spatial domain view point, facilitating selective access and manipulation of individual content.

8.3.1.1 Proposed Approach for FMV Segmentation

Based on the 3-D segmentation algorithm discussed in Chapter 6, we propose a novel unsupervised methodology for segmentation of FMV that partitions its constituents by identifying homogeneous sub-volumes/content within the data treated as a 3-D spatiotemporal volume with multiple attributes. The fundamental principle that achieves FMV content segregation via segmentation in 3-D is illustrated in Fig. 8.1, where we show sample frames of a vehicle undergoing linear motion (from left to right) at three different time instants (t_1, t_2 and t_3). On the far right, a 3-D cube involving two spatial dimensions (x, y) and one temporal dimension (t) of this spatiotemporal volume is depicted. It is imperative to observe that within this 3-D volume the object of interest undergoing linear motion forms a 3-D sub-volume, which when separated out as a single 3-D entity (also known as an object tunnel), is equivalent to partitioning

the object across the individual frames of the entire FMV sequence. Additionally, since each pixel of the spatiotemporal volume can be associated with a multi-dimensional vector of attribute (spectral, gradient, motion, textural) values, an FMV cube can be generically referred to as a 3-D vector field.

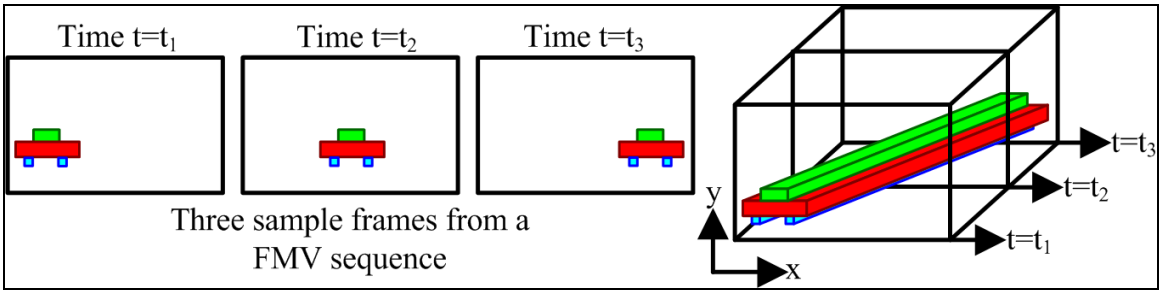

Fig. 8.1: Fundamental principle of operation of the proposed FMV segmentation algorithm.

Our approach is commenced by subjecting the input FMV to a 3-D gradient detection method that determines the magnitude of color changes across the volume. The computed gradient is utilized to guide a volume growing procedure, initiated at spatio-temporal locations with small gradient magnitudes and concluded at locations with large gradient magnitudes, to yield an initial set of homogeneous sub-volumes. These partitions are further refined by integrating them with an entropy-based texture descriptor as well as color and gradient features in a multivariate volume merging procedure that fuses sub-volumes with similar attributes, to yield the final segmentation.

8.3.1.2 Preliminary Results

In Fig. 8.2, the preliminary segmentation outcomes of our approach on five video streams in comparison to the state-of-the-art Efficient Hierarchical Graph-based Video Segmentation (EHGVS) algorithm developed by Grundmann et al. [Grundmann et al., 2010] are shown. Moreover, the results of the proposed algorithm were obtained by setting the minimum seed volume (MSV), maximum color distance (ΔE_{max}) for parent-child similarity, and maximum number of output segments (S) to 250, 5, and 50 respectively. On the other hand, the EHGVS method was run with default parametric settings from an open source web link provided by the Georgia Institute of Technology School of interactive computing.

In the results of the 'toy cars' sequence (first three rows of Fig. 8.2), observe that the EHGVS technique over-segments the background due to slight illumination variations (see Fig. 8.2(b)). However, the proposed framework enables efficient segmentation of the background, as seen in Fig. 8.2(c). Additionally, this video also demonstrates a challenging scenario wherein a foreground object (green car) undergoes various degrees of partial occlusion as the video progresses. In these circumstances, observe that the EHGVS algorithm fails to maintain the same segment labels assigned to the car across the frames, due to occlusion. Conversely, our algorithm that employs our unique 3-D statistical volume merging procedure in which several similar regions though independent are assigned the same segment label helps

Fig. 8.2: Segmentation results: (a) input video frames, (b) EHGVS algorithm, and (c) proposed algorithm.

incapacitate the occlusion problem. The segmented outputs of the 'claire' sequence are shown in rows 4 and 5 of Fig. 8.2. In the results of this video, observe that the EHGVS protocol over-segments the hair and face region due to texture disparities, while the blue jacket is segmented as two regions due to them being spatially independent, as depicted in Fig. 8.2(b). In contrast, the proposed scheme that utilizes an entropy-based texture descriptor successfully delineates the hair and face regions in the presence of the texture variations (Fig. 8.2(c)). Additionally, our multivariate volume merging approach facilitates segmentation of the jacket as one coherent region. The aforementioned observations can also be made in the outcomes of the 'car phone' sequence, portrayed in rows 6 and 7 of Fig. 8.2. Furthermore, this sequence is another example that exhibits the occlusion problem discussed previously. Observe that in frame 107, presented in row 7 of Fig. 8.2, the sky and tree regions in the background are partially occluded by the interior of the car. While the EHGVS procedure treats these regions as two different segments and even merges the tree and sky regions as one segment to the left of the occlusion, the proposed algorithm provides a more satisfactory result for reasons provided previously. The segmentation results of the 'hall' and 'mobile' sequences are displayed in rows 8 through 11 of Fig. 8.2. These are fairly complex sequences with contrasting illumination and texture variations across the scene. The outcomes for these sequences indicate our algorithm's efficiency in achieving acceptable segmentations, with superior qualitative performance in comparison to the EHGVS method.

Fig. 8.3 shows a preliminary segmentation outcome of the proposed methodology across several frames of a synthetic FMV airborne sequence, simulated via the Digital Imaging and Remote Sensing Image Generation (DIRSIG) model. These results show algorithm's potential to various FMVs acquires from airborne sensors and UAVs. It is imperative to note that while the obtained partition maps in Figs. 8.2 and 8.3 seem satisfactory with respect to the input FMV content major enhancements and future research are necessitated to improve the robustness of the algorithm. Among these is the incorporation of optical flow and motion information into the proposed segmentation framework, to achieve enhanced segmentation outcomes especially in scenarios comprising motion blur, camera panning, or multiple moving objects that may be spectrally and texturally similar. This work also requires a way of a quantitatively evaluating the performance for FMV data to facilitate objective judgment of the obtained results.

8.3.2 Integration of Segmentation and Evaluation

The bounty of segmentation methods has resulted in an increased requirement of supervised/unsupervised performance evaluation methodologies. Although, the development of evaluation strategies has been the focus of several research undertakings, it has not been proportionate to the number of algorithms established to perform the task itself. This provides an opportunity for improvement as well as opens new

avenues of research such as integration of segmentation and evaluation. It is our conviction that since segmentation is an ill posed problem with no perfect solution, the ultimate algorithm will be one that eventually performs segmentation and evaluation in an adaptive feedback mechanism, in order to define an optimal or a generally acceptable solution.

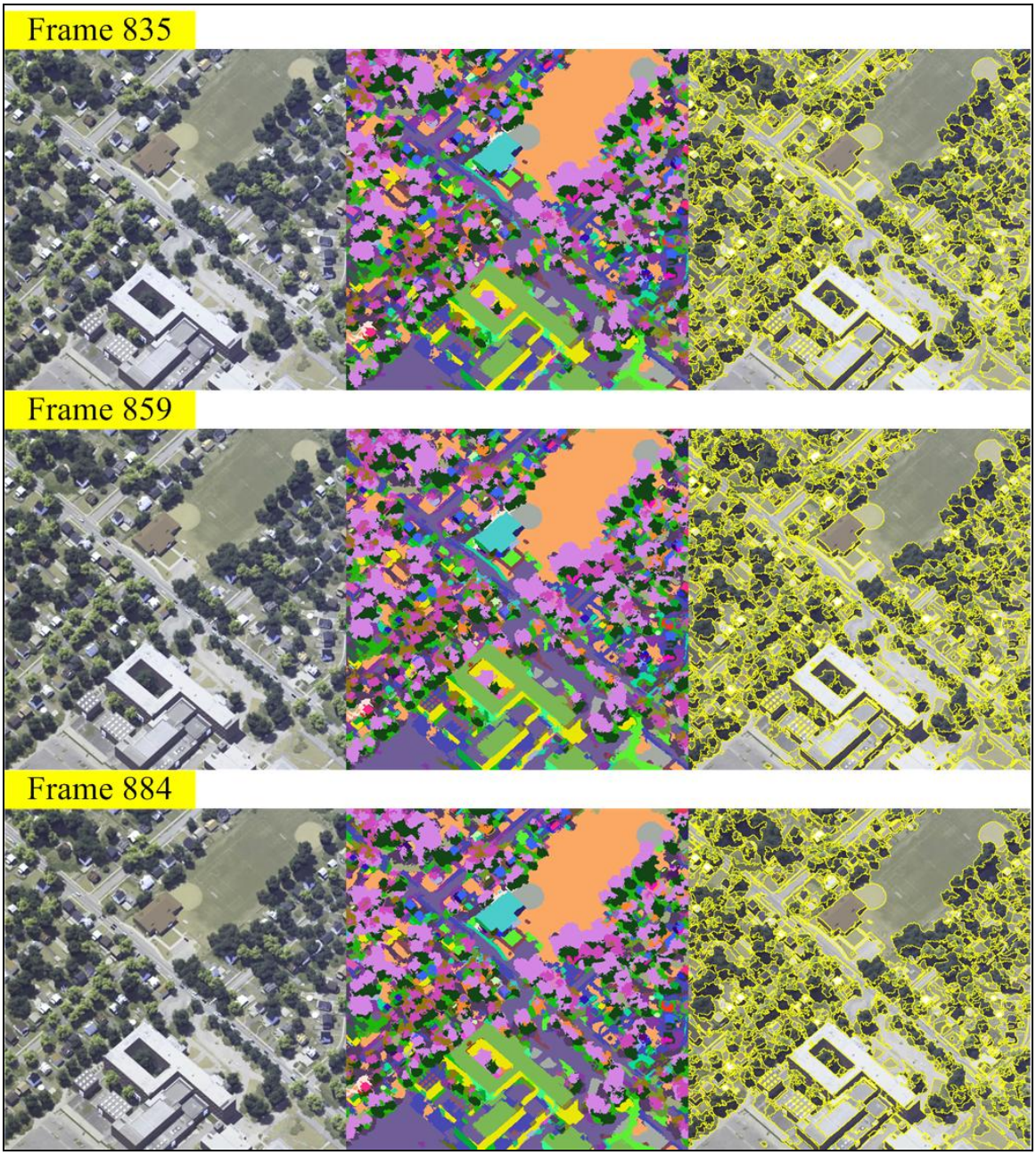

Fig. 8.3: Segmentation results acquired from a synthetic FMV airborne sequence simulated using DIRSIG.

Appendix A

Principle Component Analysis

Principle Component Analysis (PCA) also known Principle Component Transformation (PCT) as is often employed with the objective of deriving an alternate N-dimensional data space in which the original data set may exist without exhibiting correlation. To this effect, if the aforementioned alternate multi-dimensional data space does exist, the covariance matrix derived for this linearly-transformed data would be diagonal. Consequently, if data points in the alternate data space were represented by y, and G is the linear transform applied to generate the alternate data space, then:

$$y = Gx \tag{A.1}$$

holds good, with the constraint that the covariance matrix of the data in y-space should essentially be diagonal. In order to facilitate the above transformation, the PCA algorithm resorts to a squared error computation which identifies orthogonal axes for dimensionality reduction by performing an Eigen-decomposition of the covariance estimate of the original input data, represented as [Duda et al., 2001]:

$$\Sigma_x = E[(x - \mu_x)(x - \mu_x)^t] \tag{A.2}$$

where μ_x is the mean vector of the pixel data set. The resulting Eigen-decomposition under the constraint that Eq. (A.1) is valid can be expressed as:

$$\Sigma_y = G\Sigma_x G^t \tag{A.3}$$

Furthermore, since the requirement on Σ_y is that it be diagonal (mentioned previously), it follows that G is the transposed matrix of eigenvectors provided that G is orthogonal, and if this is true then Σ_y, is a diagonal matrix of the eigenvalues of Σ_x. The magnitude of an eigenvalue indicates the energy residing in the data along the component of the data parallel to the associated Eigen vector. In performing the aforementioned transformation, the larger eignvalues identify the basis components whose average contribution to $x - \mu_x$ in the squared-error sense is greater than those with smaller eigenvalues. Hence the effective dimensionality of the data can be estimated by counting the number of significantly non-zero eigenvalues and retaining only those having significant magnitude. The result is a lower-dimensional multivariate random vector that conveys most of the energy present in the original data set.

Appendix B

Quantitative Evaluation of Segmentation Methods Using Multiple Ground Truths

To objectively measure the quality of our segmentation results, we have selected a recently proposed generic technique of evaluating segmentation correctness, referred to as the Normalized Probabilistic Rand (NPR) index [Unnikrishnan et al., 2007], designed such that: 1) it does not yield cases where the evaluation produces a high value in spite of the automatic segmentation result being nowhere closely similar to any one of its corresponding ground truths (non-degeneracy), 2) no assumptions are made about label assignments and region sizes, 3) it penalizes the evaluation score when the automatic segmentation fails to distinguish between regions that humans can distinctly identify and facilitates for lesser penalty in regions that are visually ambiguous (adaptive accommodation of label refinement), and 4) it facilitates comparison amongst multiple ground truths of the same image as well as of different images. The following sub-sections briefly discuss the mathematical preliminaries required for implementing the NPR evaluation methodology.

B.1 Rand Index (R)

The Rand Index, first instituted by William Rand [Unnikrishnan and Hebert, 2005], facilitates the comparison of two arbitrary segmentations utilizing pair wise label relationships. It is defined as the ratio of number of pixel pairs that share the same label relationship in two segmentations, and is represented as:

$$R(S,S') = \frac{1}{\binom{N}{2}} \sum_{\substack{i,j \\ i \neq j}} \left[\mathrm{I}(l_i = l_j \wedge l_i' = l_j') + \mathrm{I}(l_i \neq l_j \wedge l_i' \neq l_j') \right] \tag{B.1}$$

Here S and S' are two segmentations of an image comprising of N pixels, with corresponding label assignments $\{l_i\}$ and $\{l_i'\}$ where $i = 1, 2, \ldots, N$. Furthermore, I is the identity function, ^ represents a logical conjunction ('AND' operation), and the denominator represents all possible unique pixel pairs in a dataset of N points. The Rand Index varies from 0 to 1, where 0 represents complete dissimilarity and 1 symbolizes that S and S' are identical. The Rand index is disadvantaged by its capability of handling only one ground truth segmentation for evaluation and its inability to accommodate adaptive label refinement.

B.2 Probabilistic Rand (PR) Index

The Probabilistic Rand Index [Unnikrishnan and Hebert, 2005] enables evaluation of segmentation correctness, taking into consideration the statistical nature of the Rand Index. The PR index allows comparison of a test segmentation result (S_{test}) to a set of multiple ground-truths ($S_1, S_2, \ldots S_K$) through a soft non-uniform weighting of pixel pairs as a function of the variability in the ground-truth set [Unnikrishnan et al., 2007]. The Probabilistic Rand (PR) Index is defined as:

$$PR(S_{test}, \{S_K\}) = \frac{1}{\binom{N}{2}} \sum_{\substack{i,j \\ i \neq j}} \left[\mathrm{I}(l_i^{S_{test}} = l_j^{S_{test}}) \mathrm{P}(\hat{l}_i = \hat{l}_j) + \mathrm{I}(l_i^{S_{test}} \neq l_j^{S_{test}}) \mathrm{P}(\hat{l}_i \neq \hat{l}_j) \right] \tag{B.2}$$

where $\{l_i^{S_{test}}\}$, $\{l_i^{S_K}\}$ respectively represent the label assignment of a pixel i (where $i = 1,2,\ldots N$) in S_{test} and the K^{th} manual segmentation (S_K), while $\hat{l}_i$ denote the set of "true labels" for a pixel x_i. In addition, $\mathrm{P}(\hat{l}_i = \hat{l}_j)$ and $\mathrm{P}(\hat{l}_i \neq \hat{l}_j)$ represent the respective probabilities of an identical or distinct label relationship between a pair of pixels x_i and x_j, defined as:

$$\mathrm{P}(\hat{l}_i = \hat{l}_j) = \frac{1}{K} \sum_{k=1}^{K} \mathrm{I}(l_i^{S_k} = l_j^{S_k}) \tag{B.3}$$

$$\mathrm{P}(\hat{l}_i \neq \hat{l}_j) = \frac{1}{K} \sum_{k=1}^{K} \mathrm{I}(l_i^{S_k} \neq l_j^{S_k}) = 1 - \mathrm{P}(\hat{l}_i = \hat{l}_j) \tag{B.4}$$

The PR Index takes the same range of values as the Rand Index, from 0 to 1 where 0 signifies complete dissimilarity and 1 represents a perfect match with ground truths. Although the PR Index helps overcome the aforementioned drawbacks of the Rand Index, it suffers from a deficiency of variation in its values over a large set of images due to its small effective range combined with the variation in its maximum value across images. Moreover the interpretation of the PR index across images is often ambiguous.

B.3 Normalized Probabilistic Rand (NPR) Index

In order to incapacitate the aforementioned shortcomings of the PR index, Unnikrishnan et al. [Unnikrishnan et al., 2007] proposed the Normalized Probabilistic Rand (NPR) Index. The NPR metric is referenced to expected value of the PR Index, and is computed utilizing the variation and randomness in the set of ground truth images, defined as:

$$NPR = \frac{PR - \mathrm{ExpectedIndex}}{\max[PR] - \mathrm{ExpectedIndex}} = \frac{PR - E[PR]}{\max[PR] - E[PR]} \tag{B.5}$$

The normalization with respect to the expected value of the PR Index results in a much higher range of values, making the NPR Index a much more robust evaluation metric. In Eq. (B.5) the maximum value of

the PR Index is chosen to be 1 ($\max[PR]=1$), and the expected value of the PR Index ($E[PR]$) is obtained as:

$$E[PR(S_{test},\{S_K\})] = \frac{1}{\binom{N}{2}_{i\neq j}} \sum_{i,j} \left[E[\mathrm{I}(l_i^{S_{test}} = l_j^{S_{test}})]\mathrm{P}(\hat{l}_i = \hat{l}_j) + E[\mathrm{I}(l_i^{S_{test}} \neq l_j^{S_{test}})]\mathrm{P}(\hat{l}_i \neq \hat{l}_j) \right] \quad \text{(B.6)}$$

To make the computation of $E[\mathrm{I}(l_i^{S_{test}} = l_j^{S_{test}})]$ meaningful Unnikrishnan et al. [Unnikrishnan et al., 2007] proposed its computation from segmentations of all images from an arbitrary database, for all unordered pixel pairs (i, j). Therefore, if Φ is the number of images in the database and K_Φ is the number of ground truths per image then $E[\mathrm{I}(l_i^{S_{test}} = l_j^{S_{test}})]$ and $E[\mathrm{I}(l_i^{S_{test}} \neq l_j^{S_{test}})]$ can be computed by:

$$E[\mathrm{I}(l_i^{S_{test}} = l_j^{S_{test}})] = \frac{1}{\Phi}\sum_{\Phi}\frac{1}{\mathrm{K}_\Phi}\sum_{k=1}^{k=\mathrm{K}_\Phi}\mathrm{I}(l_i^{S_k^\Phi} = l_j^{S_k^\Phi}) \quad \text{(B.7)}$$

$$E[\mathrm{I}(l_i^{S_{test}} \neq l_j^{S_{test}})] = 1 - \frac{1}{\Phi}\sum_{\Phi}\frac{1}{\mathrm{K}_\Phi}\sum_{k=1}^{k=\mathrm{K}_\Phi}\mathrm{I}(l_i^{S_k^\Phi} = l_j^{S_k^\Phi}) \quad \text{(B.8)}$$

where, $E[\mathrm{I}(l_i^{S_{test}} = l_j^{S_{test}})]$ signifies that $E[PR(S_{test},\{S_K\})]$ is a weighted sum of $PR(S_K^\Phi,\{S_K\})$. The NPR index computed in Eq. (B.5) can possess both positive and negative values, where negative values occur when the PR index is lower than its expected value for a given segmentation, signifying a poor result, while positive values that are significantly greater than zero (with a maximum value of 1) are considered useful segmentations.

Bibliography

Chapter 1

[1] [Martin et al., 2001] D. Martin, C. Fowlkes, D. Tal, and J. Malik, “A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics,” *IEEE International Conference on Computer Vision*, Vol. 2, pp. 416–423, 2001.

Chapter 2

[2] [Fu et al., 1981] K. S. Fu and J. K. Mui, “A Survey on Image Segmentation," *Elsevier Pattern Recognition*, Vol. 13, No. 1, pp. 3-16, 1981.

[3] [Pal et al., 1993] N. R. Pal and S. K. Pal, “A Review on Image Segmentation Techniques," *Elsevier Pattern Recognition*, Vol. 26, No. 9, pp. 1277-1294, 1993.

[4] [Lucchese et al., 2001] L. Lucchese and S. K. Mitra, “Color Image Segmentation: A State-of-the-Art Survey,” *Proceedings of Indian National Science Academy*, Vol. 67, No. 2, pp. 207-221, 2001.

[5] [Cheng et al., 2001] H. D. Cheng, X. H. Jiang, Y. Sun and J. L. Wang, “Color Image Segmentation: Advances & Prospects,” *Pattern Recognition Society*, Vol. 34, No. 12, pp. 2259-2281, 2001.

[6] [Green et al., 2002] P. Green and L. MacDonald, Color Engineering, *John Wiley and Sons Ltd*, 2002.

[7] [Arbeláez et al., 2006] P. A. Arbeláez and L. D. Cohen, “A Metric Approach to Vector-Valued Image Segmentation,” *International Journal of computer Vision*, Vol. 69, No. 1, pp. 119–126, 2006.

[8] [Wang et al., 2009] J. Wang, L. Ju, and X. Wang, “An Edge-Weighted Centroidal Voronoi Tessellation Model for Image Segmentation,” *IEEE Transactions on Image Processing*, Vol. 18, No. 8, pp. 1844–1858, 2009.

[9] [McQueen, 1967] J. McQueen, “Some Methods for Classification and Analysis of Multivariate Observations," *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, Vol. 1, pp. 281-296, 1967.

[10] [Kanungo et al., 2007] T. Kanungo, D. M. Mount, N. S. Netanyahu, C. D. Piatko, R. Silverman, and A. Y. Wu, “An Efficient k-Means Clustering Algorithm: Analysis and Implementation,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 7, pp. 881–892, 2007.

[11] [Chen et al., 2005] J. Chen, T. N. Pappas, A. Mojsilovic, and B. E. Rogowitz, "Adaptive Perceptual Color-Texture Image Segmentation," *IEEE Transactions on Image Processing*, Vol. 14, No. 10, pp. 1524–1536, 2005.

[12] [Mignotte, 2008] M. Mignotte, "Segmentation by Fusion of Histogram-Based K-Means Clusters in Different Color Spaces," *IEEE Transactions on Image Processing,* Vol. 17, No. 5, pp. 780–787, 2008.

[13] [Mignotte, 2011] M. Mignotte, "A De-texturing and Spatially Constrained K-means Approach for Image Segmentation," *Elsevier Pattern Recognition Letters,* Vol. 32, No. 2, pp. 359–367, 2011.

[14] [Comaniciu et al., 2002] D. Comaniciu, and P. Meer, "Mean shift: A Robust Approach Toward Feature Space Analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 5, pp. 603–619, 2002.

[15] [Fukunaga et al., 1975] K. Fukunaga, and L. D. Hosteler, "The Estimation of the Gradient of a Density Function, with Applications in Pattern -Recognition," *IEEE Transactions on Information Theory,* Vol. 21, No. 1, pp. 32–40, 1975.

[16] [Cheng, 1995] Y. Cheng, "Mean Shift, Mode Seeking, and Clustering," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 17, No. 8, pp. 790–799, 1995.

[17] [Christoudias et al., 2002] C. M. Christoudias, B. Georgescu, and P. Meer, "Synergism in Low-Level Vision," *16th IEEE International Conference on Pattern Recognition,* Vol. 4, pp. 150-155, 2002.

[18] [Hong et al., 2002] Y. Hong, J. Yi, and D. Zhao, "Improved Mean Shift Segmentation Approach for Natural Images," *Elsevier Applied Mathematics and Computation*, Vol. 24, No. 5, pp. 603–619, 2002.

[19] [Ozden et al., 2007] M. Ozden, and E. Polat, "A Color Image Segmentation Approach for Content-Based Image Retrieval," *Elsevier Pattern Recognition,* Vol. 40, No. 4, pp. 1318–1325, 2007.

[20] [Kohonen, 1990] T. Kohonen, "The Self-Organizing Map," *Proceeding of the IEEE,* Vol. 78, No. 9, pp. 1464–1480, 1990.

[21] [Huang et al., 2002] H. Y. Huang, Y. S. Chen, and W. H. Hsu, "Color Image Segmentation Using a Self-Organizing Map Algorithm," *Journal of Electronic Imaging*, Vol. 11, No. 2, 136-148, 2002.

[22] [Ong et al., 2002] S. H. Ong, N. C. Yeo, K. H. Lee, Y. V. Venkatesh, and D. M. Cao, "Segmentation of Color Images Using a Two-Stage Self-Organizing Network," *Elsevier Image and Vision Computing*, Vol. 20, No. 4, pp. 279–289, 2002.

[23] [Li et al., 2003] N. Li and Y. F. Li, "Feature Encoding for Unsupervised Segmentation of Color Images" *IEEE Transactions on Systems, Man and Cybernetics-Part B: Cybernetics*, Vol. 33, No. 3, pp. 438–447, 2003.

[24] [Dong et al., 2005] G. Dong, and M. Xie, "Color Clustering and Learning for Image Segmentation Based on Neural Networks" *IEEE Transactions on Neural Networks*, Vol. 16, No. 4, pp. 925–936, 2005.

[25] [Yeo et al., 2005] N. C. Yeo, K. H. Lee, Y. V. Venkatesh, and S. H. Ong, "Colour Image Segmentation Using the Self-Organizing Map and Adaptive Resonance Theory," *Elsevier Image and Vision Computing*, Vol. 23, No. 12, pp. 1060–1079, 2005.

[26] [Araújo et al., 2009] A. R.F. Araújo, and D. C. Costa, "Local Adaptive Receptive Field Self-Organizing Map for Image Color Segmentation," *Elsevier Image and Vision Computing*, Vol. 27, No. 9, pp. 1229–1239, 2009.

[27] [Frisch, 2006] A. S. Frisch, "Unsupervised Construction of Fuzzy Measures Through Self-Organizing Feature Maps and its Application in Color Image Segmentation," *Elsevier International Journal of Approximate Reasoning*, Vol. 41, No. 1, pp. 23–42, 2006.

[28] [Ilea et al., 2008] D. E. Ilea and P. F. Whelan, "CTex—An Adaptive Unsupervised Segmentation Algorithm Based on Color-Texture Coherence," *IEEE Transactions on Image Processing*, Vol. 17, No. 10, pp. 1926–1939, 2008.

[29] [Dunn, 1974] J. C. Dunn, "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters", *Journal of Cybernetics* Vol. 3, No. 3, pp. 32-57, 1974

[30] [Yang et al., 2002] J. F. Yang, S. S. Hao, and P. C. Chung, "Color Image Segmentation Using Fuzzy C-Means and Eigenspace Projections," *Elsevier Signal Processing*, Vol. 82, No. 3, pp. 461–472, 2002.

[31] [Liew et al., 2005] A. W. C. Liew, H. Yan, and N. F. Law, "Image Segmentation Based on Adaptive Cluster Prototype Estimation," *IEEE Transactions on Fuzzy Systems*, Vol. 13, No. 4, pp. 444–453, 2005.

[32] [Chen et al., 2005] Y. S. Chen, B. T. Chen, and W. H. Hsu, "Efficient Fuzzy C-Means Clustering for Image Data," *Journal of Electronic Imaging*, Vol. 14, No. 1, 013017 (1-12), 2005.

[33] [Hung et al., 2008] W. L. Hung, M. S. Yang, D. H. Chen, "Bootstrapping Approach to Feature-Weight Selection in Fuzzy C-Means Algorithms with an Application in Color Image Segmentation," *Elsevier Pattern Recognition Letters*, Vol. 29, No. 9, pp. 1317–1325, 2008.

[34] [Tziakos et al., 2009] I. Tziakos, C. Theoharatos, N. A. Laskaris and G. Economou, "Color Image Segmentation Using Laplacian Eigenmaps," *Journal of Electronic Imaging*, Vol. 18, No. 2, 023004 (1-10), 2009.

[35] [Krinidis et al., 2010] S. Krinidis and V. Chatzis, "A Robust Fuzzy Local Information C-Means Clustering Algorithm," *IEEE Transactions on Image Processing,* Vol. 19, No. 5, pp. 1328–1337, 2010.

[36] [Wang et al., 2010] X. Y. Wang, and J. Bu, "A Fast and Robust Image Segmentation Using FCM with Spatial Information," *Elsevier Digital Signal Processing,* Vol. 20, No. 4, pp. 1173-1182, 2010.

[37] [Yu et al., 2010] Z. Yu, O. C. Au, R. Zou, W. Yu, and J. Tian, "An Adaptive Unsupervised Approach Toward Pixel Clustering and Color Image Segmentation," *Elsevier Pattern Recognition,* Vol. 43, No. 5, pp. 1889–1906, 2010.

[38] [Veenman et al., 2003] C. J. Veenman, M. J. T. Reinders, and E. Backer, "A Cellular Coevolutionary Algorithm for Image Segmentation," *IEEE Transactions on Image Processing,* Vol. 12, No. 3, pp. 304–316, 2003.

[39] [Allili et al., 2010] M. S. Allili, D. Ziou, N. Bouguila, and S. Boutemedjet, "Image and Video Segmentation by Combining Unsupervised Generalized Gaussian Mixture Modeling and Feature Selection," *IEEE Transactions on Circuit and Systems for Video Technology,* Vol. 20, No. 10, pp. 1373–1377, 2010.

[40] [Jeon et al., 2006] B. K. Jeon, Y. B. Jung, and K. S. Hong, "Image segmentation by Unsupervised Sparse Clustering," *Elsevier Pattern Recognition Letters,* Vol. 27, No. 14, pp. 1650–1664, 2006.

[41] [Aghbari et al., 2006] Z. A. Aghbari, and R. A. Haj, "Hill-manipulation: An Effective Algorithm for Color Image Segmentation," *Elsevier Image and Vision Computing,* Vol. 24, No. 8, pp. 894–903, 2006.

[42] [Ma et al., 2007] Y. Ma, H. Derksen, W. Hong, and J. Wright, "Segmentation of Multivariate Mixed Data via Lossy Data Coding and Compression," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 29, No. 9, pp. 1546–1562, 2007.

[43] [Yang et al., 2008] A. Y. Yang, J. Wright, Y. Ma, and S. S. Sastry, "Unsupervised Segmentation of Natural Images via Lossy Data Compression," *Elsevier Computer Vision and Image Understanding,* Vol. 110, No. 2, pp. 212–225, 2008.

[44] [Huang et al., 2011] R. Huang, N. Sang, D. Luo, and Q. Tang, "Image Segmentation via Coherent Clustering in L*a*b* Color Space," *Elsevier Pattern Recognition Letters,* Vol. 32, No. 7, pp. 891–902, 2011.

[45] [Sezgin et al., 2004] M. Sezgin, and B. Sankur, "Survey Over Image Thresholding Techniques and Quantitative Performance Evaluation," *Journal of Electronic Imaging,* Vol. 13, No. 1, 146–165, 2004.

[46] [Barthel, 2006] K. U. Barthel , "3D-Data Representation with ImageJ", *ImageJ User and Developer Conference*, 2006

[47] [Kurugollu et al., 2001] F. Kurugollu, B. Sankur, and A. E. Harmanci, "Color Image Segmentation Using Histogram Multithresholding and Fusion," *Elsevier Image and Vision Computing*, Vol. 19, No. 13, pp. 915-928, 2001.

[48] [Cheng et al., 2002] H. D. Cheng, X. H. Jiang, J. Wang, "Color Image Segmentation based on Homogram Thresholding and Region Merging," *Elsevier Pattern Recognition*, Vol. 35, No. 2, pp. 373–393, 2002.

[49] [Mushrif et al., 2008] M. M. Mushrif, and A. K. Ray, "Color Image Segmentation: Rough-set Theoretic Approach," *Elsevier Pattern Recognition Letters*, Vol. 29, No. 4, pp. 483-493, 2008.

[50] [Mushrif et al., 2009] M. M. Mushrif, and A. K. Ray, "A-IFS Histon Based Multithresholding Algorithm for Color Image Segmentation," *IEEE Signal Processing Letters*, Vol. 16, No. 3, pp. 168-171, 2009.

[51] [Manay et al., 2003] S. Manay, and A. Yezzi, "Anti-Geometric Diffusion for Adaptive Thresholding and Fast Segmentation," *IEEE Transactions on Image Processing*, Vol. 12, No. 11, pp. 1310–1323, 2003.

[52] [Fan et al., 2001] J. Fan, D. K. Y. Yau, A. K. Elmagarmid, and W. G. Aref, "Automatic Image Segmentation by Integrating Color-Edge Extraction and Seeded Region Growing," *IEEE Transactions on Image Processing*, Vol. 10, No. 10, pp. 1454–1466, 2001.

[53] [Wan et al., 2003] S. Y. Wan, and W. E. Higgins, "Symmetric Region Growing," *IEEE Transactions on Image Processing*, Vol. 12, No. 9, pp. 1007–1015, 2003.

[54] [Fondón et al., 2006] I. Fondón, C. Serrano, and B. Acha, "Color-Texture Image Segmentation based on Multistep Region Growing," *SPIE Optical Engineering*, Vol. 45, No. 5, pp. 057002 (1-9), 2006.

[55] [Qin et al., 2010] A. K. Qin, and D. A. Clausi, "Multivariate Image Segmentation Using Semantic Region Growing With Adaptive Edge Penalty," *IEEE Transactions on Image Processing*, Vol. 19, No. 8, pp. 2157–2170, 2010.

[56] [Devaux et al., 2001] J. C. Devaux, P. Gouton, and F. Truchetet, "Karhunen-Loe`ve Transform Applied to Region-Based Segmentation of Color Aerial Images," *SPIE Optical Engineering*, Vol. 40, No. 7, pp. 1302–1308, 2001.

[57] [Chen et al., 2004] H. C. Chen, W. J. Chien, and S. J. Wang, "Contrast-Based Color Image Segmentation," *IEEE Signal Processing Letters*, Vol. 11, No. 7, pp. 641-644, 2004.

[58] [Nock et al., 2004] R. Nock and F. Nielsen, "Statistical Region Merging," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 11, pp. 1452-1458, 2004.

[59] [Nock et al., 2005] R. Nock and F. Nielsen, "Semi-supervised Statistical Region Refinement for Color Image Segmentation," *Elsevier Pattern Recognition*, Vol. 38, No. 6, pp. 835–846, 2005.

[60] [Kim, 2005] C. Kim, "Segmenting a Low-Depth-of-Field Image Using Morphological Filters and Region Merging," *IEEE Transactions on Image Processing*, Vol. 14, No. 10, pp. 1503–1511, 2005.

[61] [Kuan et al., 2008] Y. H. Kuan, C. M. Kuo, and N. C. Yang, "Color-Based Image Salient Region Segmentation Using Novel Region Merging Strategy," *IEEE Transactions on Multimedia*, Vol. 10, No. 5, pp. 832–845, 2008.

[62] [Liu et al., 2011] Z. Liu, L. Shen, and Z. Zhang, "Unsupervised Image Segmentation based on Analysis of Binary Partition Tree for Salient Object Extraction," *Elsevier Signal Processing*, Vol. 91, No. 2, pp. 290–299, 2011.

[63] [Tan et al., 2009] Z. Tan, and N. H. C. Yung, "Merging Toward Natural Clusters," *SPIE Optical Engineering*, Vol. 48, No. 7, pp. 077202 (1-14), 2009.

[64] [Calderero et al., 2010] F. Calderero, and F. Marques, "Region Merging Techniques Using Information Theory Statistical Measures," *IEEE Transactions on Image Processing*, Vol. 19, No. 6, pp. 1567–1586, 2010.

[65] [Deng et al., 2001] Y. Deng and B. S. Manjunath, "Unsupervised Segmentation of Color-Texture Regions in Images and Video," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 8, pp. 800–810, 2001.

[66] [Wang et al., 2006] Y. Wang, J. Yang, and N. Peng, "Unsupervised Color–Texture Segmentation based on Soft Criterion with Adaptive Mean-Shift Clustering," *Elsevier Pattern Recognition Letters*, Vol. 27, No. 5, pp. 386–392, 2006.

[67] [Wang et al., 2006] Y. G. Wang, J. Yang, and Y. C. Chang, "Color–Texture Image Segmentation by Integrating Directional Operators into JSEG Method," *Elsevier Pattern Recognition Letters*, Vol. 27, No. 16, pp. 1983–1990, 2006.

[68] [Cheng, 2003] S. C. Cheng, "Region-Growing Approach to Colour Segmentation using 3-D Clustering and Relaxation Labelling," *Proceedings of IEEE Vision, Image and Signal Processing*, Vol. 150, No. 4, pp. 270–276, 2003.

[69] [Shih et al., 2005] F. Y. Shih, and S. Cheng, "Automatic Seeded Region Growing for Color Image Segmentation," *Elsevier Image and Vision Computing*, Vol. 23, No. 10, pp. 877-886, 2005.

[70] [He et al., 2007] Y. He, Y. Luo, and D. Hu, "Automatic Seeded Region Growing based on Gradient Vector Flow for Color Image Segmentation," *SPIE Optical Engineering*, Vol. 46, No. 4, pp. 047003 (1-7), 2007.

[71] [Balasubramanian et al., 2008] G. Balasubramanian, E. Saber, V. Misic, E. Peskin and M. Shaw, "Unsupervised Color Image Segmentation Using a Dynamic Color Gradient Thresholding Algorithm", *21st SPIE/IS&T: Human Vision and Electronic Imaging Symposium*, Vol. 6806, pp. 68061H (1-9), 2008.

[72] [Ugarriza et al., 2009] L. G. Ugarriza, E. Saber, S. R. Vantaram, V. Amuso, M. Shaw, and R. Bhaskar, "Automatic Image Segmentation by Dynamic Region Growth and Multiresolution Merging", *IEEE Transactions on Image Processing*, Vol. 18, No. 10, pp. 2275-2288, 2009.

[73] [Vantaram et al., 2010] S. R. Vantaram, E.Saber, S. A. Dianat, M. Shaw, and R. Bhaskar, "Multiresolution Adaptive and Progressive Gradient-based color image SEGmentation", *Journal of Electronic Imaging*, Vol. 19, No. 1, pp. 013001 (1-21), 2010.

[74] [Krinidis et al., 2009] M. Krinidis and I. Pitas, "Color Texture Segmentation Based on the Modal Energy of Deformable Surfaces", *IEEE Transactions on Image Processing*, Vol. 18, No. 7, pp. 1613-1622, 2009.

[75] [Shih et al., 2010] F. Y. Shih, and S. Cheng, "Unsupervised Colour Image Segmentation using Dual-Tree Complex Wavelet Transform," *Elsevier Computer Vision and Image Understanding*, Vol. 114, No. 7, pp. 813-826, 2010.

[76] [Panagiotakis et al., 2011] C. Panagiotakis, I. Grinias and G. Tziritas, "Natural Image Segmentation based on Tree Equipartition, Bayesian Flooding and Region Merging", *IEEE Transactions on Image Processing*, Vol. 20, No. 8, pp. 2276-2287, 2011.

[77] [Sclaroff et al., 2001] S. Sclaroff, and L. Liu, "Deformable Shape Detection and Description via Model-Based Region Grouping," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 5, pp. 475–489, 2001.

[78] [Gevers, 2002] T. Gevers, "Adaptive Image Segmentation by Combining Photometric Invariant Region and Edge Information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 6, pp. 848–852, 2002.

[79] [Navon et al., 2005] E. Navon, O. Miller, and A. Averbuch, "Color Image Segmentation based on Adaptive Local Thresholds," *Elsevier Image and Vision Computing*, Vol. 23, No. 1, pp. 69-85, 2005.

[80] [Luo et al., 2006] Q. Luo, and T. M. Khoshgoftaar, "Unsupervised Multiscale Color Image Segmentation based on MDL Principle," *IEEE Transactions on Image Processing*, Vol. 15, No. 9, pp. 2755–2761, 2006.

[81] [Prasad et al., 2006] L. Prasad, and A. N. Skourikhine, "Vectorized Image Segmentation via Trixel Agglomeration," *Elsevier Pattern Recognition*, Vol. 39, No. 4, pp. 501–514, 2006.

[82] [Lézoray et al., 2009] O. Lézoray, and C. Charrier, "Color Image Segmentation Using Morphological Clustering and Fusion with Automatic Scale Selection," *Elsevier Pattern Recognition Letters,* Vol. 30, No. 4, pp. 397–406, 2009.

[83] [Wan et al., 2011] T. Wan, N. Canagarajah, and A. Achim, "Segmentation of Noisy Colour Images using Cauchy Distribution in the Complex Wavelet Domain," *IET Image Processing,* Vol. 5, No. 2, pp. 159-170, 2011.

[84] [Sobieranski et al., 2011] A. C. Sobieranski, E. Comunello, and A. V. Wangenheim, "Learning a Nonlinear Distance Metric for Supervised Region-Merging Image Segmentation," *Elsevier Computer Vision and Image Understanding*, Vol. 115, No. 2, pp. 127-139, 2011.

[85] [Kass et al., 1988] M. Kass, A. Witkin, and D. Terzopoulos, "Snakes: Active Contour Models," *International Journal of computer Vision,* Vol. 1, No. 4, pp. 321-331, 1988.

[86] [Dumitras et al., 2001] A. Dumitras, and A. N. Venetsanopoulos, "Angular Map-Driven Snakes With Application to Object Shape Description in Color Images," *IEEE Transactions on Image Processing*, Vol. 10, No. 12, pp. 1851–1859, 2001.

[87] [Precioso et al., 2005] F. Precioso, M. Barlaud, T. Blu, and M. Unser, "Robust Real-Time Segmentation of Images and Videos Using a Smooth-Spline Snake-based algorithm," *IEEE Transactions on Image Processing*, Vol. 14, No. 7, pp. 910–924, 2005.

[88] [Ozertem et al., 2007] U. Ozertem, and D. Erdogmus, "Nonparametric Snakes," *IEEE Transactions on Image Processing*, Vol. 16, No. 9, pp. 2361–2368, 2007.

[89] [Lankton et al., 2008] S. Lankton, and A. Tannenbaum, "Localizing Region-Based Active Contours," *IEEE Transactions on Image Processing*, Vol. 17, No. 11, pp. 2029–2039, 2008.

[90] [Malladi et al., 2001] R. Malladi, J. A. Sethian, and B. C. Vemuri, "Shape Modeling with Front Propagation: A Level Set Approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 17, No. 2, pp. 158–175, 1995.

[91] [Osher et al., 1988] S. Osher, and J.A. Sethian, "Fronts Propagating with Curvature-Dependent Speed: Algorithms based on Hamilton-Jacobi Formulations", *Journal of Computational Physics*, Vol. 79, No. 1, 12-49, 1988.

[92] [Sumengen, 2005] B. Sumengen, "A Matlab Toolbox Implementing Level Set Methods," [Online] Available: http://barissumengen.com/level_set_methods/, 2005.

[93] [Brox et al., 2006] T. Brox and J. Weickert, "Level Set Segmentation With Multiple Regions", *IEEE Transactions on Image Processing*, Vol. 15, No. 10, pp. 3213-3218, 2006.

[94] [Michailovich et al., 2007] O. Michailovich, Y. Rathi, and A. Tannenbaum, "Image Segmentation Using Active Contours Driven by the Bhattacharyya Gradient Flow", *IEEE Transactions on Image Processing*, Vol. 16, No. 11, pp. 2787-2801, 2007.

[95] [Ayed et al., 2008] I. B. Ayed, and A. Mitiche, "A Region Merging Prior for Variational Level Set Image Segmentation", *IEEE Transactions on Image Processing*, Vol. 17, No. 12, pp. 2301-2311, 2008.

[96] [Bertelli et al., 2008] L. Bertelli, B. Sumengen, B. S. Manjunath, and F. Gibou, "A Variational Framework for Multiregion Pairwise-Similarity-Based Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30, No. 8, pp. 1400–1414, 2008.

[97] [Xie et al., 2008] X. Xie, and M. Mirmehdi, "MAC: Magnetostatic Active Contour Model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30, No. 4, pp. 632–646, 2008.

[98] [Xie et al., 2010] X. Xie, "Active Contouring Based on Gradient Vector Interaction and Constrained Level Set Diffusion," *IEEE Transactions on Image Processing*, Vol. 19, No. 1, pp. 154–164, 2010.

[99] [Krinidis et al., 2009] S. Krinidis and V. Chatzis, "Fuzzy Energy-Based Active Contours," *IEEE Transactions on Image Processing*, Vol. 18, No. 12, pp. 2747–2755, 2009.

[100] [C. Li et al., 2010] C. Li, C. Xu, C. Gui, and M. D. Fox, "Distance Regularized Level Set Evolution and Its Application to Image Segmentation," *IEEE Transactions on Image Processing*, Vol. 19, No. 12, pp. 3243–3254, 2010.

[101] [Salah et al., 2010] M. B. Salah, A. Mitiche, and I. B. Ayed, "Effective Level Set Image Segmentation with a Kernel Induced Data Term" *IEEE Transactions on Image Processing*, Vol. 19, No. 1, pp. 220–232, 2010.

[102] [Karoui et al., 2010] I. Karoui, R. Fablet, J. M. Boucher, and J. M. Augustin, "Variational Region-Based Segmentation Using Multiple Texture Statistics" *IEEE Transactions on Image Processing*, Vol. 19, No. 12, pp. 3146–3156, 2010.

[103] [Ghosh et al., 2010] P. Ghosh, L. Bertelli, B. Sumengen, and B. S. Manjunath, "A Nonconservative Flow Field for Robust Variational Image Segmentation" *IEEE Transactions on Image Processing*, Vol. 19, No. 2, pp. 478–490, 2010.

[104] [Wang et al., 2011] W. Wang and R. Chung, "Image Segmentation by Optimizing a Homogeneity Measure in a Variational Framework," *Journal of Electronic Imaging*, Vol. 13, No. 1, 013009 (1-11), 2011.

[105] [Goldenberg et al., 2001] R. Goldenberg, R. Kimmel, E. Rivlin, and M. Rudzsky, "Fast Geodesic Active Contours" *IEEE Transactions on Image Processing*, Vol. 10, No. 10, pp. 1467–1475, 2001.

[106] [Sagiv et al., 2001] C. Sagiv, N. A. Sochen, and Y. Y. Zeevi, "Integrated Active Contours for Texture Segmentation" *IEEE Transactions on Image Processing*, Vol. 15, No. 6, pp. 1633–1646, 2006.

[107] [Chan et al., 2001] T. F. Chan, and L. A. Vese, "Active Contours Without Edges" *IEEE Transactions on Image Processing*, Vol. 10, No. 2, pp. 266–277, 2001.

[108] [Tsai et al., 2001] A. Tsai, A. Yezzi Jr., and A. S. Willsky, "Curve Evolution Implementation of the Mumford–Shah Functional for Image Segmentation, Denoising, Interpolation, and Magnification" *IEEE Transactions on Image Processing*, Vol. 10, No. 8, pp. 1169–1186, 2001.

[109] [Heiler et al., 2005] M. Heiler and C. Schnorr, "Natural Image Statistics for Natural Image Segmentation," *International Journal of computer Vision*, Vol. 63, No. 1, pp. 5–19, 2005.

[110] [Gao et al., 2005] S. Gao, and T. D. Bui, "Image Segmentation and Selective Smoothing by Using Mumford–Shah Model" *IEEE Transactions on Image Processing*, Vol. 14, No. 10, pp. 1537–1549, 2005.

[111] [Mukherjee, 2002] J. Mukherjee, "MRF Clustering for Segmentation of Color Images," *Elsevier Pattern Recognition Letters*, Vol. 23, No. 8, pp. 917–929, 2002.

[112] [Gao et al., 2002] J. Gao, J. Zhang, and M. G. Fleming, "Novel Technique for Multiresolution Color Image Segmentation," *SPIE Optical Engineering*, Vol. 41, No. 3, pp. 608-614, 2002.

[113] [Luo et al., 2003] J. Luo, and C. Guo, "Perceptual Grouping of Segmented Regions in Color Images," *Elsevier Pattern Recognition*, Vol. 36, No. 12, pp. 2781–2792, 2003.

[114] [Deng et al., 2004] H. Deng, D. A. Clausi, "Unsupervised Image Segmentation using a Simple MRF Model with a New Implementation Scheme," *Elsevier Pattern Recognition*, Vol. 37, No. 12, pp. 2323–2335, 2004.

[115] [Tab et al., 2006] F. A. Tab, G. Naghdy, and A. Mertins, "Scalable Multiresolution Color Image Segmentation," *Elsevier Signal Processing*, Vol. 86, No. 7, pp. 1670–1687, 2006.

[116] [Xia et al., 2006] Y. Xia, D. Feng, and R. Zhao, "Adaptive Segmentation of Textured Images by Using the Coupled Markov Random Field Model" *IEEE Transactions on Image Processing*, Vol. 15, No. 11, pp. 3559–3566, 2006.

[117] [Kato et al., 2006] Z. Kato, and T. C. Pong, "A Markov Random Field Image Segmentation Model for Color Textured Images," *Elsevier Image and Vision Computing*, Vol. 24, No. 10, pp. 1103–1114, 2006.

[118] [Kato, 2008] Z. Kato, "Segmentation of Color Images via Reversible Jump MCMC Sampling," *Elsevier Image and Vision Computing*, Vol. 26, No. 3, pp. 361–371, 2008.

[119] [Diplaros et al., 2007] A. Diplaros, N. Vlassis, and T. Gevers, "A Spatially Constrained Generative Model and an EM Algorithm for Image Segmentation" *IEEE Transactions on Neural Networks*, Vol. 18, No. 3, pp. 798–808, 2007.

[120] [Nikou et al., 2007] C. Nikou, N. P. Galatsanos, and A. C. Likas, "A Class-Adaptive Spatially Variant Mixture Model for Image Segmentation" *IEEE Transactions on Image Processing*, Vol. 16, No. 4, pp. 1121–1130, 2007.

[121] [Nikou et al., 2010] C. Nikou, A. C. Likas, and N. P. Galatsanos, "A Bayesian Framework for Image Segmentation With Spatially Varying Mixtures" *IEEE Transactions on Image Processing*, Vol. 19, No. 9, pp. 2278–2289, 2010.

[122] [Mignotte, 2010] M. Mignotte, "A Label Field Fusion Bayesian Model and Its Penalized Maximum Rand Estimator for Image Segmentation" *IEEE Transactions on Image Processing*, Vol. 19, No. 6, pp. 1610–1624, 2010.

[123] [Chen et al., 2010] S. Chen, L. Cao, Y. Wang, J. Liu, and X. Tang, "Image Segmentation by MAP-ML Estimations" *IEEE Transactions on Image Processing*, Vol. 19, No. 9, pp. 2254–2264, 2010.

[124] [Vantaram et al., 2011] S. R. Vantaram, and E.Saber, "An Adaptive Bayesian Clustering and Multivariate Region Merging-based Technique for Efficient Segmentation of Color Images", *36th IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Prague, Czech Republic, pp. 1077–1080, 2011.

[125] [Zhang et al., 2010] L. Zhang, and Q. Ji, "Image Segmentation with a Unified Graphical Model," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 8, pp. 1406–1425, 2010.

[126] [Lee et al., 2010] S. H. Lee, H. I. Koo, and N. I. Cho, "Image segmentation algorithms based on the machine learning of features," *Elsevier Pattern Recognition Letters*, Vol. 31, No. 14, pp. 2325–2336, 2010.

[127] [Carson et al., 2002] C. Carson, S. Belongie, H. Greenspan, and J. Malik, "Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 8, pp. 1026–1038, 2002.

[128] [Khan et al., 2009] J. F. Khan, R. R. Adhami, and S. M. A. Bhuiyan, "A Customized Gabor filter for Unsupervised Color Image Segmentation," *Elsevier Image and Vision Computing*, Vol. 27, No. 4, pp. 489–501, 2009.

[129] [Park et al., 2009] J. H. Park, G. S. Lee, and S. Y. Park, "Color Image Segmentation using Adaptive Mean Shift and Statistical Model-based Methods," *Elsevier Computers and Mathematics with Applications*, Vol. 57, No. 6, pp. 970–980, 2009.

[130] [Greggio et al., 2011] N. Greggio, A. Bernardino, C. Laschi, P. Dario, and J. S. Victor, "Fast estimation of Gaussian Mixture Models for Image Segmentation," *Machine Vision and Applications*, pp. 1–17, 2011.

[131] [Liu et al., 2011] Z. Liu, Y. Q. Song, J. M. Chen, C. H. Xie, and F. Zhu, "Color Image Segmentation using Nonparametric Mixture Models with Multivariate Orthogonal Polynomials," *Neural Computing and Applications*, pp. 1–11, 2011.

[132] [Arbeláez et al., 2004] P. A. Arbeláez, and L. D. Cohen, "Energy Partitions and Image Segmentation," *Journal of Mathematical Imaging and Vision*, Vol. 20, No. 1-2, pp. 43–57, 2004.

[133] [Jermyn et al., 2001] I. H. Jermyn, and H. Ishikawa, "Globally Optimal Regions and Boundaries as Minimum Ratio Weight Cycles," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 10, pp. 1075–1088, 2001.

[134] [Schoenemann et al., 2011] T. Schoenemann, S. Masnou, and D. Cremers, "The Elastic Ratio: Introducing Curvature Into Ratio-Based Image Segmentation" *IEEE Transactions on Image Processing*, Vol. 20, No. 9, pp. 2565–2581, 2011.

[135] [Unger et al., 2008] M. Unger, T. Pock, W. Trobin, D. Cremers, and H. Bischof, "TVSeg - Interactive Total Variation Based Image Segmentation" *British Machine Vision Conference (BMVC)*, Vol. 1, pp. 1–10, 2008.

[136] [Donoser et al., 2009] M. Donoser, M. Urschler, M. Hirzer and H. Bischof, "Saliency Driven Total Variation Segmentation," *12th IEEE International Conference on Computer Vision (ICCV)*, Vol. 1, pp. 817–824, 2009.

[137] [Pock et al., 2009] T. Pock, A. Chambolle, D. Cremers, and H. Bischof, "A Convex Relaxation Approach for Computing Minimal Partitions," *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 810–817, 2009.

[138] [Klodt et al., 2011] M. Klodt and D. Cremers, "A Convex Framework for Image Segmentation with Moment Constraints," *13th IEEE International Conference on Computer Vision (ICCV)*, Vol. 1, pp. 2236–2243, 2011.

[139] [Brown et al., 2012] E. Brown, T. Chan, and X. Bresson, "Completely Convex Formulation of the Chan-Vese Image Segmentation Model," *International Journal of computer Vision*, Vol. 98, No. 1, pp. 103-121, 2012.

[140] [Allili et al., 2006] M. S. Allili, D. Ziou, "Globally Adaptive Region Information for Automatic Color–Texture Image Segmentation," *Elsevier Pattern Recognition Letters*, Vol. 28, No. 15, pp. 1946–1956, 2007.

[141] [Adam et al., 2009] A. Adam, R. Kimmel, and E. Rivlin, "On Scene Segmentation and Histograms-Based Curve Evolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 9, pp. 1708–1714, 2009.

[142] [Kokkinos et al., 2009] I. Kokkinos, G. Evangelopoulos, and P. Maragos, "Texture Analysis and Segmentation Using Modulation Features, Generative Models, and Weighted Curve Evolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 1, pp. 142–157, 2009.

[143] [Li et al., 2009] Z. Li and J. Fan, "Stochastic Contour Approach for Automatic Image Segmentation," *Journal of Electronic Imaging*, Vol. 18, No. 4, 043004 (1-12), 2009.

[144] [Tu et al., 2002] Z. Tu, and S. C. Zhu, "Image Segmentation by Data-Driven Markov Chain Monte Carlo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 5, pp. 657–673, 2002.

[145] [Ko et al., 2005] B. C. Ko, and H. Byun, "FRIP: A Region-Based Image Retrieval Tool Using Automatic Image Segmentation and Stepwise Boolean AND Matching," *IEEE Transactions on Multimedia*, Vol. 7, No. 1, pp. 105–113, 2005.

[146] [Orbanz et al., 2008] P. Orbanz, and J. M. Buhmann, "Nonparametric Bayesian Image Segmentation," *International Journal of computer Vision*, Vol. 77, No. 1-3, pp. 25–45, 2008.

[147] [Scarpa et al., 2010] G. Scarpa, R. Gaetano, M. Haindl, and J. Zerubia, "Hierarchical Multiple Markov Chain Model for Unsupervised Texture Segmentation" *IEEE Transactions on Image Processing*, Vol. 18, No. 8, pp. 1830–1843, 2009.

[148] [Zhang et al., 2011] L. Zhang, and Q. Ji, "A Bayesian Network Model for Automatic and Interactive Image Segmentation" *IEEE Transactions on Image Processing*, Vol. 20, No. 9, pp. 2582–2593, 2011.

[149] [Petrovic et al., 2004] A. Petrovic, O. D. Escoda, and P. Vandergheynst, "Multiresolution Segmentation of Natural Images: From Linear to Nonlinear Scale-Space Representations" *IEEE Transactions on Image Processing*, Vol. 13, No. 8, pp. 1104–1114, 2004.

[150] [Dong et al., 2006] X. Dong, and I. Pollak, "Multiscale Segmentation With Vector-Valued Nonlinear Diffusions on Arbitrary Graphs" *IEEE Transactions on Image Processing*, Vol. 15, No. 7, pp. 1993–2005, 2006.

[151] [Sofou et al., 2008] A. Sofou and P. Maragos, "Generalized Flooding and Multicue PDE-Based Image Segmentation" *IEEE Transactions on Image Processing*, Vol. 17, No. 3, pp. 364–376, 2008.

[152] [Kay et al., 2009] D. A. Kay and A. Tomasi, "Color Image Segmentation by the Vector-Valued Allen–Cahn Phase-Field Model: A Multigrid Solution" *IEEE Transactions on Image Processing*, Vol. 18, No. 10, pp. 2330–2339, 2009.

[153] [Wu et al., 1993] Z. Wu and R. Leahy, "An Optimal Graph Theoretic Approach to Data Clustering: Theory and Its Application to Image Segmentation" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 11, pp. 1101–1113, 1993.

[154] [Shi et al., 2000] J. Shi and J. Malik, "Normalized Cuts and Image Segmentation" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, pp. 888–905, 2000.

[155] [Malik et al., 2001] J. Malik, S. Belongie, T. Leung, and J. Shi, "Contour and Texture Analysis for Image Segmentation," *International Journal of computer Vision*, Vol. 43, No. 1, pp. 7–27, 2001.

[156] [Tao et al., 2007] W. Tao, H. Jin, and Y. Zhang, "Color Image Segmentation Based on Mean Shift and Normalized Cuts" *IEEE Transactions on Systems, Man and Cybernetics-Part B: Cybernetics*, Vol. 37, No. 5, pp. 1382–1389, 2007.

[157] [Sarkar et al., 2000] S. Sarkar, and P. Soundararajan, "Supervised Learning of Large Perceptual Organization: Graph Spectral Partitioning and Learning Automata" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 5, pp. 504–525, 2000.

[158] [Gdalyahu et al., 2001] Y. Gdalyahu, D. Weinshall, and M. Werman, "Self-Organization in Vision: Stochastic Clustering for Image Segmentation, Perceptual Grouping, and Image Database Organization" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, No. 10, pp. 1053–1074, 2001.

[159] [Wang et al., 2001] S. Wang, and J. M. Siskind, "Image Segmentation with Minimum Mean Cut" *8th IEEE International Conference on Computer Vision (ICCV)*, Vol. 1, pp. 517–524, 2001.

[160] [Wang et al., 2003] S. Wang, and J. M. Siskind, "Image Segmentation with Ratio Cut" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 6, pp. 675–690, 2003.

[161] [Kim et al., 2009] J. S. Kim, and K. S. Hong, "Color Texture Segmentation using Unsupervised Graph Cuts," *Elsevier Pattern Recognition*, Vol. 42, No. 5, pp. 735–750, 2009.

[162] [Boykov et al., 2001] Y. Y. Boykov, and M. P. Jolly, "Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in N-D Images," *8th IEEE International Conference on Computer Vision (ICCV)*, Vol. 1, pp. 105–112, 2001.

[163] [Boykov et al., 2006] Y. Y. Boykov, and G. F. Lea, "Graph Cuts and Efficient N-D Image Segmentation," *International Journal of computer Vision,* Vol. 70, No. 2, pp. 109–131, 2006.

[164] [Boykov et al., 2001] Y. Y. Boykov and V. Kolmogorov, "An Experimental Comparison of Min-cut/Max-flow Algorithms for Energy Minimization in Vision," *3rd International Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR), Lecture Notes in Computer Science*, Vol. 2134, pp. 359–374, 2001.

[165] [Boykov et al., 2004] Y. Y. Boykov, and V. Kolmogorov, "An Experimental Comparison of Min-cut/Max-flow Algorithms for Energy Minimization in Vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 26, No. 9, pp. 1124–1137, 2004.

[166] [Gulshan et al., 2010] V. Gulshan, C. Rother, A. Criminisi, A. Blake and A. Zisserman, "Geodesic Star Convexity for Interactive Image Segmentation," [Online] Available: http://www.robots.ox.ac.uk/~vgg/research/iseg/, 2010.

[167] [Rother et al., 2004] C. Rother, V. Kolmogorov, and A. Blake, "GrabCut: Interactive Foreground Extraction Using Iterated Graph Cuts," *ACM Transactions on Graphics (SIGGRAPH)*, Vol. 23, pp. 309–314, 2004.

[168] [Han et al., 2009] S. Han, W. Tao, D. Wang, X. C. Tai, and X. Wu, "Image Segmentation Based on GrabCut Framework Integrating Multiscale Nonlinear Structure Tensor" *IEEE Transactions on Image Processing*, Vol. 18, No. 10, pp. 2289–2302, 2009.

[169] [Li et al., 2004] Y. Li, J. Sun, C. K. Tang, and H. Y. Shum, "Lazy Snapping," *ACM Transactions on Graphics (SIGGRAPH)*, Vol. 23, pp. 303-308, 2004,

[170] [Yang et al., 2007] Q. Yang, X. Tang, C. Wang, Z. Ye, and M Chen, "Progressive Cut: An Image Cutout Algorithm that Models User Intentions," *IEEE Multimedia*, Vol. 14, No. 3, pp. 56-66, 2007,

[171] [Feng et al., 2010] W. Feng, J. Jia, and Z. Q. Liu, "Self-Validated Labeling of Markov Random Fields for Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 10, pp. 1871–1887, 2010.

[172] [Felzenszwalb et al., 2004] P. F. Felzenszwalb, and D. P. Huttenlocher, "Efficient graph-based image segmentation," *International Journal of Computer Vision*, Vol. 59, No. 2, pp. 167–181, 2004.

[173] [Ding et al., 2008] J. Ding, R. Ma, and S. Chen, "A Scale-Based Connected Coherence Tree Algorithm for Image Segmentation" *IEEE Transactions on Image Processing*, Vol. 17, No. 2, pp. 204–216, 2008.

[174] [Dupuis et al., 2006] A. Dupuis, and P. Vasseur, "Image Segmentation by Cue Selection and Integration," *Elsevier Image and Vision Computing*, Vol. 24, No. 10, pp. 1053-1064, 2006.

[175] [Grady et al., 2006] L. Grady, and E. L. Schwartz, "Isoperimetric Graph Partitioning for Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 3, pp. 469–475, 2006.

[176] [Grady, 2006] L. Grady, "Random Walks for Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 11, pp. 1768–1783, 2006.

[177] [Yang et al., 2010] W. Yang, J. Cai, J. Zheng, and J. Luo, "User-Friendly Interactive Image Segmentation Through Unified Combinatorial User Inputs" *IEEE Transactions on Image Processing*, Vol. 19, No. 9, pp. 2470–2479, 2010.

[178] [Xiang et al., 2011] S. Xiang, C. Pan, F. Nie, and C. Zhang, "Interactive Image Segmentation With Multiple Linear Reconstructions in Windows" *IEEE Transactions on Multimedia*, Vol. 13, No. 2, pp. 342–352, 2011.

[179] [Liu et al., 2010] G. Liu, Z. Lin, Y. Yu, and X. Tang, "Unsupervised Object Segmentation with a Hybrid Graph Model (HGM)," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 5, pp. 910–924, 2010.

[180] [Ding et al., 2010] L. Ding, and A. Yilmaz, "Interactive Image Segmentation Using Probabilistic Hypergraphs," *Elsevier Pattern Recognition*, Vol. 43, No. 5, pp. 1863–1873, 2010.

[181] [Tao et al., 2010] W. Tao, F. Chang, L. Liu, H. Jin, and T. Wang, "Interactively Multiphase Image Segmentation based on Variational Formulation and Graph Cuts," *Elsevier Pattern Recognition*, Vol. 43, No. 10, pp. 3208–3218, 2010.

[182] [Salah et al., 2011] M. B. Salah, A. Mitiche, and I. B. Ayed, "Multiregion Image Segmentation by Parametric Kernel Graph Cuts" *IEEE Transactions on Image Processing*, Vol. 20, No. 2, pp. 545–557, 2011.

[183] [Makrogiannis et al., 2005] S. Makrogiannis, G. Economou, and S. Fotopoulos, "A Region Dissimilarity Relation That Combines Feature-Space and Spatial Information for Color Image Segmentation" *IEEE Transactions on Systems, Man and Cybernetics-Part B: Cybernetics*, Vol. 35, No. 1, pp. 44–53, 2005.

[184] [Makrogiannis et al., 2005] S. Makrogiannis, G. Economou, S. Fotopoulos, and N. G. Bourbakis, "Segmentation of Color Images Using Multiscale Clustering and Graph Theoretic Region Synthesis" *IEEE Transactions on Systems, Man and Cybernetics-Part A: Systems and Humans*, Vol. 35, No. 2, pp. 224–238, 2005.

[185] [Sumengen et al., 2006] B. Sumengen, and B.S. Manjunath, "Graph Partitioning Active Contours (GPAC) for Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 28, No. 4, pp. 509–521, 2006.

[186] [Wangenheim et al., 2007] A. V. Wangenheim, R. F. Bertoldi, D. D. Abdala, and M. M. Richter, "Color Image Segmentation Guided by a Color Gradient Network," *Elsevier Pattern Recognition Letters*, Vol. 28, No. 13, pp. 1795–1803, 2007.

[187] [Wangenheim et al., 2009] A. V. Wangenheim, R. F. Bertoldi, D. D. Abdala, A. Sobieranski, L. Coser, X. Jiang, M. M. Richter, L. Priese, and F. Schmitt, "Color Image Segmentation Using an Enhanced Gradient Network Method," *Elsevier Pattern Recognition Letters*, Vol. 30, No. 15, pp. 1404–1412, 2009.

[188] [Kumar et al., 2010] M. P. Kumar, P.H.S. Torr, and A. Zisserman, "OBJCUT: Efficient Segmentation Using Top-Down and Bottom-Up Cues," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 32, No. 3, pp. 530–545, 2010.

[189] [Couprie et al., 2011] C. Couprie, L. Grady, L. Najman, and H. Talbot, "Power Watershed: A Unifying Graph-Based Optimization Framework," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 7, pp. 1384–1399, 2011.

[190] [Gonzalez et al., 2008] R. C. Gonzalez and R. E. Woods, Digital Image Processing 3rd Edition, *Pearson/Prentice Hall Ltd.*, 2008.

[191] [Gao et al., 2001] H. Gao, W. C. Siu, and C. H. Hou, "Improved Techniques for Automatic Image Segmentation" *IEEE Transactions on Circuits and systems for Video Technology*, Vol. 11, No. 12, pp. 1273–1280, 2001.

[192] [Gao et al., 2006] H. Gao, W. Lin, P. Xue, and W. C. Siu, "Marker-based Image Segmentation Relying on Disjoint Set Union," *Elsevier Signal Processing: Image communication*, Vol. 21, No. 2, pp. 100–112, 2006.

[193] [Hill et al., 2003] P. R. Hill, C. N. Canagarajah, and D. R. Bull, "Image Segmentation Using a Texture Gradient Based Watershed Transform" *IEEE Transactions on Image Processing*, Vol. 12, No. 12, pp. 1618–1633, 2003.

[194] [Nguyen et al., 2003] H. T. Nguyen, M. Worring, and R. Boomgaard, "Watersnakes: Energy-Driven Watershed Segmentation" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 25, No. 3, pp. 330–342, 2003.

[195] [Vanhamel et al., 2003] I. Vanhamel, I. Pratikakis, and H. Sahli, "Multiscale Gradient Watersheds of Color Images" *IEEE Transactions on Image Processing*, Vol. 12, No. 6, pp. 617–626, 2003.

[196] [Kim et al., 2003] J. B. Kim, and H. J. Kim, "Multiresolution-based Watersheds for Efficient Image Segmentation," *Elsevier Pattern Recognition Letters*, Vol. 24, No. 1-3, pp. 473–488, 2003.

[197] [Jung, 2007] C. R. Jung, "Unsupervised Multiscale Segmentation of Color Images," *Elsevier Pattern Recognition Letters*, Vol. 28, No. 4, pp. 523–533, 2007.

[198] [Arbeláez et al., 2011] P. Arbeláez, M. Maire, C. Fowlkes, and J. Malik, "Contour Detection and Hierarchical Image Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 5, pp. 898–916, 2011.

[199] [O'Callaghan et al., 2005] R. J. O'Callaghan and D. R. Bull, "Combined Morphological-Spectral Unsupervised Image Segmentation" *IEEE Transactions on Image Processing*, Vol. 14, No. 1, pp. 49–62, 2005.

[200] [Makrogiannis et al., 2005] S. Makrogiannis, I. Vanhamel, S. Fotopoulos, H. Sahli and J. P. H. Cornelis, "Watershed-based Multiscale Segmentation Method for Color Images using Automated Scale Selection," *Journal of Electronic Imaging*, Vol. 14, No. 3, 033007 (1-16), 2009.

[201] [Cheng et al., 2003] H. D. Cheng, and J. Li, "Fuzzy Homogeneity and Scale-Space Approach to Color Image Segmentation," *Elsevier Pattern Recognition*, Vol. 36, No. 7, pp. 1545–1562, 2003.

[202] [Chaabane et al., 2010] S. B. Chaabane, M. Sayadi, F. Fnaiech, and E. Brassart, "Colour Image Segmentation Using Homogeneity Method and Data Fusion Techniques," *EURASIP Journal on Advances in Signal Processing*, Vol. 2010, No. 367297, pp. 1–11, 2010.

[203] [Choy et al., 2011] S. K. Choy, M. L. Tang, and C. S. Tong, "Image Segmentation Using Fuzzy Region Competition and Spatial/Frequency Information" *IEEE Transactions on Image Processing*, Vol. 20, No. 6, pp. 1473–1484, 2011.

[204] [Protiere et al., 2007] A. Protiere and G. Sapiro, "Interactive Image Segmentation via Adaptive Weighted Distances" *IEEE Transactions on Image Processing*, Vol. 16, No. 4, pp. 1046–1057, 2007.

[205] [Xiang et al., 2009] S. Xiang, F. Nie, C. Zhang, and C. Zhang, "Interactive Natural Image Segmentation via Spline Regression" *IEEE Transactions on Image Processing*, Vol. 18, No. 7, pp. 1623–1632, 2009.

[206] [Bai et al., 2009] X. Bai, and G. Sapiro, "Geodesic Matting: A Framework for Fast Interactive Image and Video Segmentation and Matting" *International Journal of Computer Vision*, Vol. 82, No. 2, pp. 113–132, 2009.

[207] [Li et al., 2010] H. Li, and C. Shen, "Interactive Color Image Segmentation with Linear Programming," *Machine Vision and Applications*, Vol. 21, No. 4, pp. 403–412, 2010.

[208] [Funt et al., 2007] L. Shi, and B. Funt, "Quaternion Color Texture Segmentation," *Elsevier Computer Vision and Image Understanding*, Vol. 107, No. 1-2, pp. 88–96, 2007.

[209] [Subakan et al., 2011] Ö. N. Subakan, and B. C. Vemuri, "A Quaternion Framework for Color Image Smoothing and Segmentation" *International Journal of Computer Vision*, Vol. 91, No. 3, pp. 233–250, 2011.

[210] [Shotton et al., 2006] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "TextonBoost: Joint Appearance, Shape and Context Modeling for Multi-Class Object Recognition and Segmentation," *European Conference on Computer Vision (ECCV)*, Vol. 1, pp. 1–15, 2006.

[211] [Shotton et al., 2008] J. Shotton, M. Johnson, and R. Cipolla, "Semantic Texton Forests for Image Categorization and Segmentation," *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–8, 2008.

[212] [Shotton et al., 2009] J. Shotton, J. Winn, C. Rother, and A. Criminisi, "TextonBoost for Image Understanding: Multi-Class Object Recognition and Segmentation by Jointly Modeling Texture, Layout, and Context" *International Journal of Computer Vision*, Vol. 81, No. 1, pp. 2–23, 2009.

[213] [Yu et al., 2012] Z. Yu, A. Li, O. Au, and C. Xu, "Bag of Textons for Image Segmentation via Soft Clustering and Convex Shift," *accepted, IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[214] [Wang et al., 2008] C. Wang, and L. Guan, "Graph Cut Video Object Segmentation using Histogram of Oriented Gradients" *IEEE International Symposium on Circuits and Systems (ISCAS)*, pp. 2590–2593, 2008.

[215] [Nammalwar et al., 2010] P. Nammalwar, O. Ghita, and P. Whelan, "A Generic Framework for Colour Texture Segmentation" *Sensor Review*, Vol. 30, No. 1, pp. 69–79, 2010.

[216] [Li et al., 2011] X. Li, H. Fan, Y. Zhao, and H. Zhang, "Graph Cuts Based Image Segmentation Using Local Color And Texture" *4th IEEE International Congress on Image and Signal Processing (CISP)*, Vol. 3, pp. 1251–1255, 2011.

[217] [Rohkohl et al., 2007] C. Rohkohl, and K. Engel, "Efficient Image Segmentation Using Pairwise Pixel Similarities" *29th Annual Symposium of the German Association for Pattern Recognition*, Vol. 4713, pp. 254–263, 2007.

[218] [Levinshtein et al., 2009] A. Levinshtein, A. Stere, K. Kutulakos, D. Fleet, S. Dickinson, and K. Siddiqi, "TurboPixels: Fast Superpixels Using Geometric Flows," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 31, No. 12, pp. 2290–2297, 2009.

[219] [Achanta et al., 2010] R. Achanta, A. Shaji, K. Smith, A. Lucchi, P. Fua, and S. Susstrunk, "SLIC Superpixels," *École Polytechnique Fédérale de Lausanne (EPFL) Technical Report 149300*, pp. 1–15, 2010.

[220] [Liu et al., 2011] M. Liu, O. Tuzel, S. Ramalingam, and R. Chellappa, "Entropy Rate Superpixel Segmentation," *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2097–2104, 2011.

[221] [Artan, 2011] Y. Artan, "Interactive Image Segmentation using Machine Learning Techniques," *IEEE Canadian Conference on Computer and Robot Vision (CRV)*, pp. 264–269, 2011.

[222] [Li et al., 2012] Z. Li, X. Wu, S. Chang, "Segmentation Using Superpixels: A Bipartite Graph Partitioning Approach," *accepted, IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[223] [Hoang et al., 2005] M. A. Hoang, J. M. Geusebroek, A. W. M. Smeulders, "Color Texture Measurement and Segmentation," *Elsevier Signal Processing*, Vol. 85, No. 2, pp. 265–275, 2005.

[224] [Vazquez et al., 2011] E. Vazquez, R. Baldrich, J. V. D. Weijer, and M. Vanrell, "Describing Reflectances for Color Segmentation Robust to Shadows, Highlights, and Textures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 33, No. 5, pp. 917–930, 2011.

[225] [Wang et al., 2011] X. Y. Wang, T. Wang, and J. Bu, "Color Image Segmentation Using Pixel wise Support Vector Machine Classification," *Elsevier Pattern Recognition*, Vol. 44, No. 4, pp. 777-787, 2011.

[226] [Yu et al., 2011] Z. Yu, H. S. Wong, G. Wen, "A Modified Support Vector Machine and its Application to Image Segmentation," *Elsevier Image and Vision Computing*, Vol. 29, No. 1, pp. 29-40, 2011.

[227] [Macaire et al., 2006] L. Macaire, N. Vandenbroucke, and J. G. Postaire, "Color Image Segmentation by Analysis of Subset Connectedness and Color Homogeneity Properties," *Elsevier Computer Vision and Image Understanding*, Vol. 102, No. 1, pp. 105–116, 2006.

[228] [Zöller et al., 2007] T. Zöller, and J. M. Buhmann, "Robust Image Segmentation Using Resampling and Shape Constraints," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 7, pp. 1147–1164, 2007.

[229] [Borenstein et al., 2008] E. Borenstein, and S. Ullman, "Combined Top-Down/Bottom-Up Segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 30, No. 12, pp. 2109–2125, 2008.

[230] [Delyon et al., 2006] G. Delyon, F. Galland, and P. Réfrégier, "Minimal Stochastic Complexity Image Partitioning With Unknown Noise Model" *IEEE Transactions on Image Processing*, Vol. 15, No. 10, pp. 3207–3212, 2006.

[231] [Bardera et al., 2009] A. Bardera, J. Rigau, I. Boada, M. Feixas, and M. Sbert, "Image Segmentation Using Information Bottleneck Method" *IEEE Transactions on Image Processing*, Vol. 18, No. 7, pp. 1601–1612, 2009.

[232] [Song et al., 2011] Y. Z. Song, B. Xiao, P. Hall, and L. Wang, "In Search of Perceptually Salient Groupings" *IEEE Transactions on Image Processing*, Vol. 20, No. 4, pp. 935–947, 2011.

[233] [Usó et al., 2011] A. M. Usó, F. Pla, and P. G. Sevilla, "Unsupervised Colour Image Segmentation by Low-Level Perceptual Grouping," *Pattern Analysis and Applications*, pp. 1-14, 2011.

[234] [Ghosh et al., 2009] S. Ghosh, J. Pfeiffer III, and J. Mulligan, "A General Framework for Reconciling Multiple Weak Segmentations of an Image" *IEEE Workshop on Applications of Computer Vision (WACV)*, pp. 1–8, 2009.

[235] [Vitaladevuni et al., 2010] S. Vitaladevuni, and R. Basri, "Co-Clustering of Image Segments Using Convex Optimization Applied to EM Neuronal Reconstruction," *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2203–2210, 2010.

[236] [Glasner et al., 2011] D. Glasner, S. Vitaladevuni, and R. Basri, "Contour-Based Joint Clustering of Multiple Segmentations," *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2385–2392, 2011.

[237] [Rother et al., 2006] C. Rother, V. Kolmogorov, T. Minka, and A. Blake, "Cosegmentation of Image Pairs by Histogram Matching —Incorporating a Global Constraint into MRFs," *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, Vol. 1, pp. 993–1000, 2006.

[238] [Cheng et al., 2007] D. Cheng, and M. Figueiredo, "Cosegmentation for Image Sequences," *14th IEEE International Conference on Image Analysis and Processing (ICIAP)*, pp. 635–640, 2007.

[239] [Hochbaum et al., 2009] D. Hochbaum, and V. Singh, "An Efficient Algorithm for Co-Segmentation," *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 269–276, 2009.

[240] [Mukherjee et al., 2009] L. Mukherjee, V. Singh, and C. Dyer, "Half-Integrality based Algorithms for Cosegmentation of Images," *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2028–2035, 2009.

[241] [Batra et al., 2010] D. Batra, A. Kowdle, and D. Parikh, "iCoseg: Interactive Cosegmentation with Intelligent Scribble Guidance," *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3169–3176, 2010.

[242] [Chang et al., 2011] K. Chang, T. Liu, and S. Lai, "From Co-saliency to Co-segmentation: An Efficient and Fully Unsupervised Energy Minimization Model," *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2129–2136, 2011.

[243] [Vicente et al., 2011] S. Vicente, C. Rother, and V. Kolmogorov, "Object Cosegmentation," *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2217–2224, 2011.

[244] [Joulin et al., 2012] A. Joulin, F. Bach, and J. Ponce, "Multi-Class Cosegmentation," *accepted, IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[245] [Kim et al., 2012] G. Kim, and E. Xing, "On Multiple Foreground Cosegmentation," *accepted, IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*, 2012.

[246] [Carleer et al., 2005] A. P. Carleer, O. Debeir, and E. Wolff, "Assessment of Very High Spatial Resolution Satellite Image Segmentations," *ASPRS Photogrammetric Engineering and Remote Sensing*, Vol. 71, No. 11, pp. 1285–1294, 2005.

[247] [Xu and Wunsch, 2005] R. Xu, and D. Wunsch, "Survey of Clustering Algorithms," *IEEE Transactions on Neural Networks*, Vol. 16, No. 3, pp. 645-678, 2005.

[248] [Bilgin et al., 2011] G. Bilgin, S. Ertürk, and T. Yıldırım, "Segmentation of Hyperspectral Images via Subtractive Clustering and Cluster Validation Using One-Class Support Vector Machines," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 49, No. 8, pp. 2936–2944, 2011.

[249] [Mercovich et al., 2011] R. A. Mercovich, A. Harkin, and D. Messinger, "Automatic Clustering of Multispectral Imagery by Maximization of the Graph Modularity," *SPIE Algorithms and Technologies for Multispecral, Hyperspectral, and Ultraspectral Imagery XVII Conference*, Vol. 8048, pp. 8048Z (1-12), 2011.

[250] [Tran et al., 2003] T. N. Tran, R. Wehrens, and L. M. C. Buydens, "SpaRef: A Clustering Algorithm for Multispectral Images," *Elsevier Analytica Chimica Acta*, Vol. 490, No. 1-2, pp. 303-312, 2003.

[251] [Lee and Lee, 2010] S. Lee and C. Lee, "Unsupervised Segmentation of Hyperspectral Images using Mean Shift Segmentation," *Proceedings of SPIE Satellite Data Compression, Communications, and Processing VI Conference*, Vol. 7810, pp. 781011 (1-6), 2010.

[252] [Paglieroni, 2003] D. W. Paglieroni, "A Self-Calibrating Multi-Band Region Growing Approach to Segmentation of Single and Multi-Band Images," *SPIE Photonics West, Optical Enginerring*, Vol. 5001, pp. 65-75, 2003.

[253] [Lee and Lee, 2008] S. Lee and C. Lee, "Unsupervised Segmentation of Hyperspectral Images," *Proceedings of SPIE Satellite Data Compression, Communication, and Processing IV Conference*, Vol. 7084, pp. 70840B (1-8), 2008.

[254] [Gorretta et al., 2009] N. Gorretta, J. Roger, G. Rabatel, C. Fiorio, and C. Lelong, "Hyperspectral Image Segmentation: the Butterfly Approach," *IEEE 1st Workshop on Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS)*, pp. 1-4, 2009.

[255] [Bouziani et al., 2010] M. Bouziani, K. Goita, and D. He, "Rule-Based Classification of a Very High Resolution Image in an Urban Environment Using Multispectral Segmentation Guided by Cartographic Data," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 48, No. 8, pp. 3198–3211, 2010.

[256] [Akçay and Aksoy, 2008] H. G. Akçay, and S. Aksoy, "Automatic Detection of Geospatial Objects Using Multiple Hierarchical Segmentations," *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 46, No. 7, pp. 2097–2111, 2008.

[257] [Li and Xiao, 2007] P. Li, and X. Xiao, "Multispectral Image Segmentation by A Multichannel Watershed-based Approach," *International Journal of Remote Sensing*, vol. 28, no. 19, 4429-4452, 2007.

[258] [Li et al., 2011] P. Li, J. Guo, B. Song, and X. Xiao, "A Multilevel Hierarchical Image Segmentation Method for Urban Impervious Surface Mapping Using Very High Resolution Imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 4, no. 1, pp. 103–116, 2011.

[259] [Tarabalka et al., 2010] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and Classification of Hyperspectral Images using Watershed Transformation", *Elsevier Pattern Recognition*, vol. 43, no. 7, pp. 2367-2379, 2010.

[260] [Anguloa et al., 2010] J. Anguloa, and S. Velasco-Foreroa, "Semi-supervised Hyperspectral Image Segmentation using Regionalized Stochastic Watershed," *SPIE Algorithms and Technologies for Multispecral, Hyperspectral, and Ultraspectral Imagery XVI Conference*, vol. 7695, pp. 76951F (1-12), 2010.

[261] [Castilla et al., 2008] G. Castilla, G. J. Hay, and J. R. Ruiz-Gallardo, "Size-Constrained Region Merging (SCRM): An Automated Delineation Tool for Assisted Photointerpretation," *ASPRS Photogrammetric Engineering and Remote Sensing*, vol. 74, no. 4, pp. 409-419, 2008.

[262] [Rand and Keenan., 2001] R. S. Rand, and D. M. Keenan, "A Gibbs-based Unsupervised Segmentation Approach to Partitioning Hyperspectral Imagery for Terrain Applications," *SPIE Algorithms and Technologies for Multispecral, Hyperspectral, and Ultraspectral Imagery VII Conference*, vol. 4381, pp. 275-285, 2001.

[263] [Farag et al., 2005] A. A. Farag, R. M. Mohamed, and A. El-Baz, "A Unified Framework for MAP Estimation in Remote Sensing Image Segmentation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 43, no. 7, pp. 1617–1634, 2005.

[264] [Li et al., 2010] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Semisupervised Hyperspectral Image Segmentation Using Multinomial Logistic Regression With Active Learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 11, pp. 4085–4098, 2010.

[265] [Li et al., 2011] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Hyperspectral Image Segmentation Using a New Bayesian Approach With Active Learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 10, pp. 3947–3960, 2011.

[266] [Li et al., 2011] J. Li, J. M. Bioucas-Dias, and A. Plaza, "Spectral–Spatial Hyperspectral Image Segmentation Using Subspace Multinomial Logistic Regression and Markov Random Fields," *in press IEEE Transactions on Geoscience and Remote Sensing*, 2011.

[267] [Borges et al., 2011] J. S. Borges, J. M. Bioucas-Dias, and A. R. S. Marcal, "Bayesian Hyperspectral Image Segmentation With Discriminative Class Learning," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 2151–2164, 2011.

[268] [Sun et al., 2011] A. Sun, Y. Tan, and J. Tian, "Hyperspectral Image Segmentation Using Spectral-Spatial Constrained Conditional Random Field," *SPIE Multispectral Image Acquisition, Processing, and Analysis Conference*, vol. 8002, pp. 800213 (1-8), 2011.

[269] [Ball and Bruce, 2007] J. E. Ball, and L. M. Bruce, "Level Set Hyperspectral Image Classification Using Best Band Analysis," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 10, pp. 3022–3027, 2007.

[270] [Duarte-Carvajalino et al., 2008] J. M. Duarte-Carvajalino, G. Sapiro, M. Vélez-Reyes, and P. E. Castillo, "Multiscale Representation and Segmentation of Hyperspectral Imagery Using Geometric Partial Differential Equations and Algebraic Multigrid Methods," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 8, pp. 2418–2434, 2008.

[271] [Mitra et al., 2004] P. Mitra, B. U. Shankar, S. K. Pal, "Segmentation of Multispectral Remote Sensing Images using Active Support Vector Machines," *Elsevier Pattern Recognition Letters*, vol. 25, no. 9, pp. 1067–1074, 2004.

[272] [Gang and Xingjian, 2008] L. Gang, and L. Xingjian, "Researches on Segmentation Method of Hyper-Spectral Remote Sensing Images Based on Probabilistic Neural Networks," *SPIE International Conference on Earth Observation Data Processing and Analysis*, vol. 7285, pp. 72851J (1-8), 2008.

[273] [Tarabalka et al., 2010] Y. Tarabalka, J. Chanussot, and J. A. Benediktsson, "Segmentation and Classification of Hyperspectral Images Using Minimum Spanning Forest Grown From Automatically Selected Markers," *IEEE Transactions on Systems, Man, and Cybernetics Part B: Cybernetics*, vol. 40, no. 5, pp. 1267–1279, 2010.

[274] [Wun and Messinger, 2011] W. Sun, and D. W. Messinger, "Tri-lateral Filter on Multispectral Imagery for Classification and Segmentation," *SPIE Algorithms and Technologies for Multispecral, Hyperspectral, and Ultraspectral Imagery XVII Conference*, vol. 8048, pp. 80480Y (1-9), 2011.

[275] [Hazel, 2000] G. G. Hazel, "Multivariate Gaussian MRF for Multispectral Scene Segmentation and Anomaly Detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 38, no. 3, pp. 1199–1211, 2000.

[276] [Hong et al., 2003] P. S. Hong, L. M. Kaplan, and M. J. T. Smith, "Hyperspectral Image Segmentation using Filter Banks for Texture Augmentation," *IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data*, pp. 254-258, 2003.

[277] [Gaetano et al., 2009] R. Gaetano, G. Scarpa, and G. Poggi, "Hierarchical Texture-Based Segmentation of Multiresolution Remote-Sensing Images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 47, no. 7, pp. 2129–2141, 2009.

[278] [Li et al., 2010] N. Li, H. Huo, and T. Fang, "A Novel Texture-Preceded Segmentation Algorithm for High-Resolution Imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 48, no. 7, pp. 2818–2828, 2010.

[279] [Pham et al., 2000] D. L. Pham, C. Xu, and J. L. Prince, "Current Methods in Medical Image Segmentation," *Annual Review of Biomedical Engineering*, vol. 2, pp. 315-337, 2000.

[280] [Bulu and Alpkocak, 2007] H. Bulu and A. Alpkocak, "Comparison of 3D Segmentation Algorithms for Medical Imaging," *IEEE International Symposium on Computer-Based Medical Systems*, vol. 15, pp. 269-274, 2007.

[281] [Withey and Koles, 2007] D. J. Withey, and Z. J. Koles, "Medical Image Segmentation: Methods and Software," *IEEE Joint Meeting of the 6th International Symposium on Noninvasive Functional Source Imaging (NFSI) of the Brain and Heart and the International Conference on Functional Biomedical Imaging (ICFBI)*, pp. 140-143, 2007.

[282] [Bomans et al., 1990] M. Bomans, K. Hohne, U. Tiede, and M. Riemer, "3-D Segmentation of MR Images of the Head for 3-D Display," *IEEE Transactions on Medical Imaging*, vol. 9, no. 2, pp. 177–183, 1990.

[283] [Yan and Karp, 1994] M. X. H. Yan and J. S. Karp, "Segmentation of 3D Brain MR Using An Adaptive K-means Clustering Algorithm," *IEEE Nuclear Science Symposium and Medical Imaging Conference*, vol. 4, pp. 1529-1533, 1994.

[284] [Pappas, 1992] T. N. Pappas, "An Adaptive Clustering Algorithm for Image Segmentation," *IEEE Transactions on Signal Processing*, vol. 40, no. 4, pp. 901–914, 1992.

[285] [Choi et al., 1997] S. M. Choi, J. E. Lee, J. Kim, and M. H. Kim, "Volumetric Object Reconstruction Using the 3D-MRF Model-Based Segmentation," *IEEE Transactions on Medical Imaging*, vol. 16, no. 6, pp. 887–892, 1997.

[286] [Justice et al., 1997] R. K. Justice, E. M. Stokely, J. S. Strobel, R. E. Ideker, and W. M. Smith, "Medical Image Segmentation using 3-D Seeded Region Growing," *Proceedings of SPIE Medical Imaging Conference*, vol. 3034, pp. 900-910, 1997.

[287] [Sijbers et al., 1997] J. Sijbers, M. Verhoye, P. Scheunders, A. Van der Linden, D. Van Dyck, and E. Raman, "Watershed-based Segmentation of 3D MR Data for Volume Quantization," *Elsevier Magnetic Resonance Imaging*, vol. 15, no. 6, pp. 679–688, 1997.

[288] [Hastreiter and Ertl, 1998] P. Hastreiter, and T. Ertl, "Fast and Interactive 3D-Segmentation of Medical Volume Data," *Proceedings of Image and Multidimensional Digital Signal Processing (IMDSP)*, pp. 41-44, 1998.

[289] [Yezzi et al., 1997] A. Yezzi, Jr., S. Kichenassamy, A. Kumar, P. Olver, and A. Tannenbaum, "A Geometric Snake Model for Segmentation of Medical Imagery," *IEEE Transactions on Medical Imaging*, vol. 16, no. 2, pp. 199–209, 1997.

[290] [Baillard et al., 2000] C. Baillard, C. Barillot, and P. Bouthemy, "Robust Adaptive Segmentation of 3D Medical Images with Level Sets," *Technical Report INRIA*, no. 1369, 2000.

[291] [Tek et al., 2002] H. Tek, M. Bergtholdt, D. Comaniciu, and J. Williams, "Segmentation of 3D Medical Structures Using Robust Ray Propagation," *Lecture Notes in Computer Science*, vol. 2488, pp. 572-579, 2002.

[292] [Freedman et al., 2005] D. Freedman, R. J. Radke, T. Zhang, Y. Jeong, D. M. Lovelock, and G. T. Y. Chen, "Model-Based Segmentation of Medical Imagery by Matching Distributions," *IEEE Transactions on Medical Imaging*, vol. 24, no. 3, pp. 281–292, 2005.

[293] [Ding et al., 2005] F. Ding, W. K. Leow, and S. Wang, "Segmentation of 3D CT Volume Images Using a Single 2D Atlas," Proceedings of the 1st International *Workshop on Computer Vision for Biomedical Image Applications*, pp. 1-10, 2005.

[294] [Yushkevich et al., 2006] P. A. Yushkevich, J. Piven, H. C. Hazlett, R. G. Smith, S. Ho, J. C. Gee, and G. Gerig, "User-Guided 3D Active Contour Segmentation of Anatomical Structures: Significantly Improved Efficiency and Reliability", *Elsevier NeuroImage*, vol. 31, no. 3, pp. 1116-1128, 2006.

[295] [Shen et al., 2009] T. Shen, H. Li, Z. Qian, and X. Huang, "Active Volume Models for 3D Medical Image Segmentation", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 707-714, 2009.

[296] [Shen et al., 2009] T. Shen, and X. Huang, "3D Medical Image Segmentation by Multiple-Surface Active Volume Models", *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 1059-1066, 2009.

[297] [Olowoyeye et al., 2009] A. Olowoyeye, M. Tuceryan and S. Fang, "Medical Volume Segmentation using Bank of Gabor Filters", *24th Annual ACM Symposium on Applied Computing (SAC)*, pp. 826-830, 2009.

[298] [Campadelli et al., 2010] P. Campadelli, E. Casiraghi, and S. Pratissoli, "A Segmentation Framework for Abdominal Organs from CT Scans", *Elsevier Artificial Intelligence in Medicine*, vol. 50, no. 1, pp. 3-11, 2010.

[299] [AlZu'bi and Amira, 2010] S. AlZu'bi, and A. Amira, "3D Medical Volume Segmentation Using Hybrid Multiresolution Statistical Approaches," *Advances in Artificial Intelligence*, vol. 2010, pp. 1–15, 2010.

[300] [Huang et al., 2010] A. Huang, R. Abugharbieh, and R. Tam, "A Novel Rotationally Invariant Region-Based Hidden Markov Model for Efficient 3-D Image Segmentation," *IEEE Transactions on Image Processing*, vol. 19, no. 10, pp. 2737-2748, 2010.

[301] [Hu et al., 2001] S. Hu, E. A. Hoffman, and J. M. Reinhardt, "Automatic Lung Segmentation for Accurate Quantitation of Volumetric X-Ray CT Images," *IEEE Transactions on Medical Imaging*, vol. 20, no. 6, pp. 490–498, 2001.

[302] [Kang et al., 2003] Y. Kang, K. Engelke, and W. A. Kalender, "A New Accurate and Precise 3-D Segmentation Method for Skeletal Structures in Volumetric CT Data," *IEEE Transactions on Medical Imaging*, vol. 22, no. 5, pp. 586–598, 2003.

[303] [Ecabert et al., 2008] O. Ecabert, J. Peters, H. Schramm, C. Lorenz, J. V. Berg, M. J. Walker, M. Vembar, M. E. Olszewski, K. Subramanyan, G. Lavi, and J. Weese, "Automatic Model-Based Segmentation of the Heart in CT Images," *IEEE Transactions on Medical Imaging*, vol. 27, no. 9, pp. 1189-1201, 2008.

[304] [Iˇsgum et al., 2009] I. Iˇsgum, M. Staring, A. Rutten, M. Prokop, M. A. Viergever, and B. V. Ginneken, "Multi-Atlas-Based Segmentation with Local Decision Fusion—Application to Cardiac and Aortic Segmentation in CT Scans," *IEEE Transactions on Medical Imaging*, vol. 28, no. 7, pp. 1000-1010, 2009.

Chapter 3

[305] [Green and MacDonald, 2002] P. Green and L. MacDonald, "Color Engineering", *John Wiley and Sons Inc.*, 2002.

[306] [Lee and Cok, 1991] H. C. Lee and D. R. Cok, "Detecting boundaries in a vector field," *IEEE Transactions on Signal Processing*, vol. 39, no. 5, pp. 1181–1194, 1991.

[307] [Haralick and Shapiro, 1992] R. M. Haralick and L. Shapiro, "Computer and Robot Vision", *Addison-Wesley*, vol. 1, pp. 28–48, 1992.

[308] [Cover and Thomas, 1991] T. Cover and J. Thomas, "Elements of Information Theory". *Wiley Interscience*, 1991.

[309] [Chou and Wu, 2003] C. H. Chou and T. L. Wu, "Embedding color watermarks in color images," *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 1, pp. 32-40, 2003.

[310] [Krzanowski, 1988] W. J. Krzanowski, "Principles of multivariate analysis: A user's perspective," *Oxford University Press*, Chapter 11, 1988

Chapter 4

[311] [Christopoulos et al., 2000] C. Christopoulos, A. Skodras, and T. Ebrahimi, "The JPEG2000 still image coding system: An overview," *IEEE Transactions on Consumer Electronics*, vol. 46, no. 4, pp. 1103-1127, 2000.

Chapter 5

[312] [Green et al., 1988] A. A. Green, M. Berman, P. Switzer, and M. D. Craig, "A Transformation for Ordering Multispectral Data in Terms of Image Quality with Implications for Noise Removal," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 26, pp. 65–74, 1988.

[313] [Lee et al., 1990] J. B. Lee, A. S. Woodyatt, and M. Berman, "Enhancement of High Spectral Resolution Remote-Sensing Data by a Noise-Adjusted Principal Components Transform", *IEEE Transactions on Geoscience and Remote Sensing*, vol. 28, no. 3, pp. 295–304, May 1990.

[314] [Prewitt, 1970] J. M. S. Prewitt, "Object enhancement and extraction," in *Picture Processing and Psychopictorics*, Academic Press, 1970.

[315] [Canny, 1986] J. Canny, "A computational approach to edge detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 679–698, 1986.

[316] [Gonzalez and Woods, 2008] R. C. Gonzalez and R. E. Woods (3rd edition), "Digital image processing," *Prentice Hall publication,* Ch. 10, 2008.

[317] [Mitchell, 1995] P. A. Mitchell, "Hyperspectral digital imagery collection experiment (HYDICE)," *Proceedings of SPIE Geographic Information Systems, Photogrammetry, and Geological/Geophysical Remote Sensing Conference*, vol. 2587, pp. 70-95, (1995).

[318] [Haralick et al., 1973] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 3, pp. 610–621, 1973.

[319] [Schott, 2007] J. R. Schott (2nd edition), "Remote sensing: The imaging chain approach," *Oxford University press,* Ch. 8, 2007.

Chapter 6

[320] [Smith, 1961] O. K. Smith, "Eigenvalues of a symmetric 3 × 3 matrix," *Communications of the Association for Computing Machinery*, Vol. 4, pp. 168, 1961.

[321] [Zwillinger, 1981] D. Zwillinger, "Standard Mathematical Tables", *CRC Press*, 26th Edition, pp. 9-11, 1981.

Chapter 7

[322] [Unnikrishnan et al., 2007] R. Unnikrishnan, C. Pantofaru, and M. Hebert, "Toward objective evaluation of image segmentation algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 6, pp. 929–944, 2007.

[323] [Folkman et al., 2001] M. A. Folkman, J. Pearlman, L. B. Liao, and P. J. Jarecke, "EO-1/Hyperion Hyperspectral Imager Design, Development, Characterization, and Calibration," *Proceedings of SPIE Hyperspectral Remote Sensing of the Land and Atmosphere Conference*, vol. 4151, pp. 40–51, 2001.

[324] [Feng et al., 1994] X. Feng, J. R. Schott, T. W. Gallagher, "Modeling and Testing of a Modular Imaging Spectrometer Instrument," *Proceedings of SPIE Infrared Imaging Systems: Design, Analysis, Modeling, and Testing V Conference*, vol. 2224, pp. 215–224, 1994.

[325] [Kunkel et al., 1991] B. Kunkel, F. Blechinger, D. Viehmann, H. Van Der Piepen, R. Doeffer, "ROSIS Imaging Spectrometer and its Potential for Ocean Parameter Measurements (Airborne and Spaceborne)," *International Journal of Remote Sensing*, vol. 12, no. 4, 753-761, 1991.

[326] [Zhang, 1996] Y. J. Zhang, "A Survey on Evaluation Methods for Image Segmentation," *Elsevier Pattern Recognition*, vol. 29, no. 8, pp. 1335-1346, 1996.

[327] [Linden et al., 1996] S. Linden, A. Janz, B. Waske, M. Eiden, and P. Hostert, "Classifying Segmented Hyperspectral Data from a Heterogeneous Urban Environment using Support Vector Machines," *J. Appl. Remote Sens.*, vol. 1, 013543, no. 8, pp. 1335-1346, 1996.

[328] [Chang and Lin, 2011] C. C. Chang, and C. J. Lin, "LIBSVM—A Library for Support Vector Machines," *ACM Trans. on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1:27-27:27, 2011.

[329] [Kennedy et al., 1989] D. N. Kennedy, P. A. Filipek, and V. S. Caviness, "Anatomic Segmentation and Volumetric Calculations in Nuclear Magnetic Resonance Imaging," *IEEE Transactions on Medical Imaging*, vol. 8, no. 1, pp. 1-7, 1989.

[330] [Filipek et al., 1994] P. A. Filipek, C. Richelme, D. N. Kennedy, and V. S. Caviness, "The Young Adult Human Brain: An MRI-based Morphometric Analysis," *Cerebral Cortex*, vol. 4, no. 4, pp. 344-360, 1994.

Chapter 8

[331] [Grundmann et al., 2010] M. Grundmann, V. Kwatra, M. Han, and I. Essa, "Efficient Hierarchical Graph-Based Video Segmentation", *IEEE International Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 2141 - 2148, 2010.

Appendix A

[332] [Duda et al., 2001] R. O. Duda, P. E. Hart, and D. G. Stork, "Pattern Classification," *John Wiley and Sons Inc.*, Chapter 3, 2001.

Appendix B

[333] [Unnikrishnan and Hebert, 2005] R. Unnikrishnan, and M. Hebert, "Measures of similarity," *IEEE Workshop on Applications of Computer Vision*, pp. 394-400, 2005.